Doctoral theses at NTNU, 2024:160

Ingrid Bondevik

Eliminating adjunct-specific conditions on movement

A case study of finite adjunct clauses in Norwegian

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

Ingrid Bondevik

Eliminating adjunct-specific conditions on movement

A case study of finite adjunct clauses in Norwegian

Thesis for the Degree of Philosophiae Doctor

Trondheim, May 2024

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Humanities
Department of Language and Literature

ISBN 978-82-326-7914-0 (printed ver.)
ISBN 978-82-326-7913-3 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:160

Printed by NTNU Grafisk senter

Abstract

The dissertation investigates variation in syntactic constraints focusing on extraction from adjunct clauses (i.e., adjunct islands) in Norwegian. The overall aim of the dissertation is to ascertain whether there is fine-grained variation in extraction patterns between types of extraction and types of adjunct clauses. The dissertation explores, through formal acceptability judgments, extraction, by way of topicalization and relativization, from three different adjunct clause types in Norwegian – conditional *om* 'if', habitual *når* 'when' and causal *fordi* 'because'.

The dissertation is composed of three papers. Papers 1 and 2 are empirically focused, while Paper 3 is theoretically focused. Paper 1 investigates whether forming a topicalization dependency into the three finite adjunct clause types mentioned above yields island effects in Norwegian. The paper finds that there is a statistically significant island effect for all three adjunct clause types. However, the size of the island effect varies as a function of adjunct clause type: the island effect for *fordi* 'because'- and *når* 'when'-clauses are classic, large island effects; while the island effect for *om* 'if'-clauses is small. Paper 2 follows up on the findings in Paper 1 by testing the same three adjunct clause types in a relativization dependency. Paper 2 shows that the findings replicate. Paper 2 also reports on a follow-up experiment investigating the small island effect size of conditional *om* 'if'-clauses. The follow-up shows that the small island effect yielded by *om* 'if'-clauses is not caused by variation between participants, items and/or order. Together, Papers 1 and 2 show that adjunct clauses are not a uniform class for extraction and that theories of extraction from adjunct clauses must be fine-grained enough to distinguish between types of finite adjunct clause types and between island effect sizes.

Paper 3 provides a theoretical explanation of the variation in island effects discovered in Papers 1 and 2. The paper proposes that the variation between adjunct clause types is syntactically conditioned. Specifically, Paper 3 proposes that the distinction between adjunct clause types is caused by syntactic differences in the left periphery of each of the embedded adjunct clauses. Second, Paper 3 suggests that the gradience in acceptability is a syntactic, representational filter on derivational output.

Together with the cover article, the papers provide evidence of fine-grained variation in syntactic constraints on movement. Specifically, the dissertation shows that the *type* of adjunct clause must be considered in theorizing on extraction from adjunct clauses. Furthermore, on a general level, and departing from much previous work on adjunct islands, the dissertation suggests that extractability varies as a function of the internal syntax of finite clause types and not as a function of the distinction between adjuncts and arguments.

Sammendrag

Denne avhandlinga undersøker variasjon i syntaktiske begrensninger. Avhandlinga fokuserer på begrensninger på utflytting fra adverbiale leddsetninger (adjunktøyer) i norsk. Det overordna målet er å bestemme hvorvidt det er finmaska variasjon i utflyttingsmønster mellom typer av flytting (f.eks. topikalisering vs. hv-flytting) og typer adverbiale leddsetninger. Avhandlinga bruker eksperimentelle aksepabilitetsvurderinger til å utforske utflytting gjennom topikalisering og relativisering fra tre ulike adverbiale leddsetninger – leddsetninger innledet av *om, når* og *fordi*.

Avhandlinga består av tre artikler. Artikkel 1 og 2 er empirisk rettet mens artikkel 3 er teoretisk rettet. Artikkel 1 undersøker hvorvidt topikalisering ut fra de tre adverbiale leddsetningene nevnt ovenfor gir øy-effekter i norsk. Artikkelen viser at det er statistisk signifikante øy-effekter ved topikalisering ut fra alle de tre leddsetningene, men at størrelsen på øy-effekten varierer basert på hvilken leddsetning det flyttes fra. Øy-effekten for *fordi*- og *når*-setninger er klassiske, store øy-effekter, mens øy-effekten for *om*-setninger er liten. Disse funnene følges opp i artikkel 2. I artikkel 2 testes de samme leddsetningene med relativisering. Resultatene med topikalisering repliseres med relativisering. Artikkel 2 følger også opp den lille øy-effekten funnet for *om*-setninger både topikalisering og relativisering. Resultatene i artikkel 2 viser at den lille øy-effekten med *om*-setninger ikke er et resultat av variasjon mellom deltagere, test-setninger eller rekkefølge. Artikkel 1 og 2 viser sammen at adverbiale leddsetninger ikke utgjør en ensartet gruppe for utflytting og at teorier om utflytting fra adverbiale leddsetninger må være finmaska nok til å kunne skille mellom typer av adverbiale leddsetninger og mellom størrelser i øy-effekter.

Artikkel 3 gir en teoretisk forklaring av variasjonen i øy-effekter avdekket i artikkel 1 og 2. Artikkel 3 foreslår at variasjon mellom adverbielle leddsetningstyper er betinget av syntaktiske forskjeller i venstreperiferien i hver av de adverbiale leddsetningene. Denne artikkelen foreslår også at mellomstore øy-effekter er en følge av et syntaktisk, representativt filter som avleser ferdigbygde strukturer før de sendes til produksjon.

Sammen med kappa presenterer artiklene bevis for at det er finmaska variasjon i syntaktiske begrensninger på utflytting. Avhandlinga viser at *typen* adverbial leddsetning må tas i betraktning ved teorisering rundt muligheten for utflytting fra adverbiale leddsetninger. Videre foreslår avhandlinga at utflytting varierer basert på den interne syntaksen til setningstyper og ikke basert på iboende forskjeller mellom adjunkter og argumenter.

Acknowledgements

Writing a PhD-dissertation is in some ways a lonely journey, but it has also allowed me to get to know so many amazing people who have helped me to learn, write and do my best research (yet) over the last four years.

First, I wish to thank the Faculty of Humanities at NTNU for believing in my project and providing funding for this dissertation. Next, I am very grateful to the faculty for assigning me such great supervisors in Professor Terje Lohndal and Associate Professor Andrew Weir. They have been amazing (scientific) role models and provided me with expert advice on everything I have ever asked about. I have learned so much from discussing small details and big-picture questions related to my project with them. I will be forever grateful that they took me under their academic wings.

This journey would not have been nearly as fun without all my colleagues and friends. My "PhD Ladies" have been incredibly important the whole way – Sunniva Briså Strætkvern, Anastasia Kobzeva, Anna Giskes and Kristin Klubbo Brodahl made it fun to come to work and actually made doing a COVID-PhD quite eventful. I especially want to thank Sunniva for being the best PhD-bestie there is – the feedback she has given on my texts and the hours we have spent writing together have been invaluable. I cannot believe how far we have come since the first day in the office when we looked at each other and said "ok, what do we do now?". It has been so nice to have a fellow adjunct-researcher in Kristin and an island-friend in Anastasia to discuss my work with. I am also grateful to the members of PRESS (Producitivty Exercise for Successful Submission) for the countless hours we have spent writing together in silence. Thanks also to Ingrid Marie Heiene, Natalie Hoage and Charlotte Sant for being great PhD-colleagues.

I also want to thank Associate Professor Dave Kush for showing me how to work hard and not give up until I have a product I can really be proud of. Dave taught me statistical thinking and to notice old ideas disguised as new proposals, and for those skills I am very grateful.

I am also very grateful to Professor Mike Putnam who accepted the invitation to be my opponent for a mock defence. Mike's insight and extensive knowledge of the field gave me a clearer direction and encouraged me to be bolder in my claims. He also made the mock defense into a great discussion that I thoroughly enjoyed.

The dissertation would not be the same without the larger community of linguists. I have been amazed by how helpful and community-oriented my fellow linguists are and I have

received much advice, help, inspiration and constructive feedback from members of AcqVa at NTNU, the EyelandsLab at NTNU, Professor Jon Sprouse, Associate Professor Natalia Mitrofanova, Myrte Vos, Associate Professor Bodo Winter at the UiT course on Linear Models, Gianluca Porta and audiences at MONS, AmLaP, NELS, workshops in Göttingen and UiA.

Finally, I want to thank my own family, my family in-law and my friends for believing in me right from the start and for distracting me with activities, phone calls and fun. I am very grateful to my volleyball team (D1 @Tbk) for providing the abosolute best diversion. There is nothing better than to throw a ball around after a long day of writing. I want to extend a special thanks to my mother-in-law, professor Rachel Jakhelln, who has been my unofficial PhD-mentor since before I even began my PhD.

Writing this thesis would have been a thousand times harder without my partner, Magnus Hemmingsen. I have really benefitted from his outside perspective on the project and on doing a PhD in general. I am sure I would not be the same researcher or human without his never-ending patience, love and support.

List of Papers

1. Bondevik, Ingrid & Kush, Dave & Lohndal, Terje. 2021. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics*, 44(3), 223-254. doi: 10.1017/S0332586520000207
2. 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), 1–41. doi: https://doi.org/10.16995/glossa.9033
3. Bondevik, Ingrid. submitted. Why adjuncts are not islands categorically: a case study of causal *fordi* 'because' and conditional *om* 'if' in Norwegian.

Notes on Papers 1 and 2

Paper 1

Paper 1 is a joint paper together with Dave Kush and Terje Lohndal. All three were involved in conceptualizing the paper. While I was in charge of data collection and data analysis, Dave Kush helped me set-up and administer the data collection tool and shared R scripts for data analysis. I have written the first draft of the paper, while both Kush and Lohndal were greatly involved in editing and framing the paper together with me. I was in charge of revisions after reviews, but all changes were thoroughly discussed with my co-authors.

Paper 2

Paper 2 is a joint paper with Terje Lohndal. I was in charge of data collection and analysis. I have written the first draft of all sections of the paper and Lohndal was involved in conceptualizing, editing and framing the paper together with me. We were both involved in designing the follow-up experiment. I was in charge of revisions after reviews and all changes were discussed with my co-author.

List of abbreviations

AC	Adjunct Condition
AdvP	Adverbial Phrase
BC	Blocking Category
BCI	Backgrounded constituents are islands
CED	Condition on Extraction Domain
CNPC	Complex Noun Phrase Constraint
COMP	Complementizer
CSC	Coordinate Structure Constraint
CP	Complementizer Phrase
DD	*differences-in-differences*
DP	Determiner Phrase
DTLC	Double Trouble Locality Condition
EEG	Electroencephalography
EST	Extended Standard Theory
FBC	Focus Background Constraint
FinP	Finiteness Phrase
fRM	featural Relativized Minimality
GG	Generative Grammar
IP	Inflection Phrase
LF	Logical Form
Mod	Modifier
MP	Minimalist Program
MSc	Mainland Scandinavian
NP	Nominal Phrase
Op	Operator
PF	Phonological Form
PIC	Phrase Impenetrability Condition
POS	Poverty of the Stimulus
PP	Prepositional Phrase
R-expression	Referring expression
Rc	Relative clause
RM	Relativized Minimality
SEGC	Single Event Grouping Condition

Spec	Specifier
SSC	Sentential Subject Constraint
Top	Topicalization
TP	Tense Phrase
UG	Universal Grammar
*v*P	Verbalizer Phrase
VP	Verbal Phrase

Table of contents

Cover article

1 Introduction

The overarching goal of this dissertation is to investigate the empirical realities of extraction from adjunct islands. Norwegian constitutes a good case study for this investigation as previous research has uncovered some surprising findings for Norwegian, as well as for the other Mainland Scandinavian languages. The traditional claim has for a long time been that adjuncts are categorical islands cross-linguistically (see Stepanov 2007; Truswell 2007, 2011; Bode 2020). This view is starting to weaken in recent years as empirical findings contradicting this view have been presented (see e.g. Chaves & Putnam 2020). The current dissertation adds to these findings by (i) showing that "adjunct" is not a uniform class for islandhood; (ii) A'-dependencies are sensitive to different adjunct clause types and; (iii) acceptability of island extraction is a gradient measure in certain cases, but not in others.

However, the new empirical findings are not adequately explained by existing theoretical approaches to adjuncts and islands. This is the secondary goal of the present dissertation. After having laid out the empirical realities of Norwegian adjunct clause extraction, I will investigate whether existing theoretical approaches to islands, and adjunct islands in particular, can explain the pattern. The dissertation provides a new proposal that adequately explains the patterns revealed for Norwegian, and which has the potential to be extended to cross-linguistic patterns of extraction from adjunct clauses.

The dissertation is comprised of three papers. The first two papers provide empirical investigations of extraction from three different finite adjunct clause types in Norwegian; conditional *om* 'if', habitual *når* 'when' and causal *fordi* 'because'. The first paper investigates these adjunct clause types in a topicalization (top)-dependency, and the second in a relative clause (*rc*)-dependency. A standardized, formal methodology for investigating islands is used for the empirical investigation. In a broad perspective, these papers provide evidence that top- and *rc*-dependencies are not distinct from each other across the different adjunct clause types, but that different adjunct clause types, though all of them show island effects to some degree, must be distinguished as they consistently yield different island effects. The third paper provides a discussion and proposal for how (and whether) these findings can be explained by existing approaches to islands. The paper finds that traditional approaches that are not adjunct-specific can provide an explanation of the findings, but only if two different locality conditions are applied in tandem. In effect, adjunct clauses that yield different island effects might be derived by different constraints. In other words, I argue that adjunct island effects cannot be captured under an umbrella constraint governing all adjunct clauses.

The cover article provides a review of the relevant literature and the methodological design and concerns of the studies in Papers 1 and 2. In addition, main findings of the three papers as a whole are synthesized and discussed, and the theoretical implications of the overall findings explored in detail. Finally, the main findings for Norwegian are investigated from a cross-linguistic perspective.

1.1 Theoretical starting point – what are islands?

Humans produce and process language seemingly effortlessly – we do not need to think about how we produce sentences, and we interpret sentences with the same ease. Yet, there are strict rules that must be followed (i.e., case-marking, order of words, verb declension, agreement etc.). In a conversation, a speaker wants to express a thought. This thought is quickly formed into a linguistic expression without the average language user having any conscious awareness of how the linguistic expression is formed.

To take the order of words as an example, as this essentially is the main topic of the current thesis, order matters greatly for how we understand a certain string of words. This is illustrated with the simple sentence pair in Table 1:

Table 1: Word order varies, interpretation also varies.

String	*Interpretation*
a. The cat chased the mouse	
b. The mouse chased the cat	

The example shows that the order of the string matters for interpretation. In English, if we want to describe a scene in which a mouse is being chased by a cat, we place *the cat* before *the mouse* (Table 1 a.). If, however, we want to describe the opposite (and quite surprising) scenario, *the mouse* precedes the cat (Table 1 b.).

Thus, we deduce that there must be rules for combining words into strings (and rules for case assignment, verb declension, agreement etc.). And furthermore, that every speaker *subconsciously* knows all the relevant rules for their native language(s). These rules make up what we refer to as 'the grammar'. The grammar makes up a mental reality that is learned from input and traditionally assumed to be aided by innate knowledge of language (*Universal Grammar* (UG), Chomsky 1965). The job of the linguist is to make this mental reality, which for Chomsky is rooted in human psychology and biology, explicit (Lasnik & Lohndal 2013: 27).

This sets the stage for the framework *(Transformational) Generative Grammar* (GG), where "generative" means "explicit" (Lasnik & Lohndal 2013: 27). Chomsky (1980: 220) writes: "When we speak of the linguist's grammar as a 'generative grammar', we mean only that it is sufficiently explicit to determine how sentences in the language are in fact characterized by the grammar". In the early stages of GG, most specifically under the *Extended Standard Theory* (EST) of the 1960s and 70s, several specific rules for building linguistic structures and specific transformational operations manipulating the basic structures were postulated. Under Principles and Parameters Theory in the 1980s, a similar system was maintained. Later, in the Minimalist Program (MP) (Chomsky 1993, 1995), however, this large set of specific rules was abandoned in favor of a simpler, minimalistic grammar. In MP the gold standard for a grammar is to be as simple and sparse in specific rules as possible. Freidin & Lasnik (2011: 4) present the MP as having an important heuristic as well as a therapeutic effect on linguistic inquiry in "limiting the hypothesis space for linguistic analysis […]. And, in particular, by prohibiting analyses that merely mirror the complexity of the data". As such, working from a minimalist perspective requires a pursuit of simpler and more minimal theories of language.[1] In "early minimalism" (Allott & Lohndal 2023) there are three basic operations; *Move, Adjoin* and *Merge* (see e.g., Adger 2003). In late minimalist accounts, there is typically only one operation; *Merge.* Merge both builds linguistic structures and manipulates them, thus, eliminating the need for the structure building rules and specific transformation rules of the EST, as well as the distinction between deep and surface structure.

I will spend some time introducing and explaining *Merge* as it is an important prerequisite for understanding the details and complexity introduced later in the cover article. In short, Merge is a general operation that combines lexical items (Hornstein, Nunes & Grohmann 2005). To build the phrase *looked at John*, there are several steps. First, as shown in (1) the preposition *at* and the nominal *John* are merged to form the PP *at John*.

[1] Related to the pursuit of simpler theories is the exploration of language as an optimal system, the idea being that the simplest system is the most optimal (Freidin & Lasnik 2011: 6). Chomsky (2000: 23) writes "[…] languages are highly imperfect in all these respects, as indeed you would expect – they have indices and bar levels, D-structures, S-structures and all kinds of relations […]. So it is no small task to demonstrate the contrary. Nevertheless, I think the contrary could well be true". Such an assertion, according to Allott & Lohndal (2023: 12), does not mean that Chomsky takes it to be true that language is optimal, but that Chomsky declares his intentions to work with and explore the idea that language is optimal (pursuit) as part of the MP.

(1) the PP *at John* is merged

{at, John}

In the next step of the derivation, the head of the phrase "projects" (i.e., is duplicated) to indicate its precedence:

(2) *at* projects

{at, {at, John}} – *at* projects

In (2), the preposition *at* takes the nominal *John* as its complement – *at* projects and labels the structure (Hornstein et al. 2005: 202). Since the preposition *at* projects, the phrase *at John* will behave as a PP. A PP like *at John*, can then be merged with a verb that takes it as its complement and forms a larger VP phrase as in (3).

(3)

a) {looked, {at, {at, John}}} – *looked* is merged with *at John*
b) {looked, {looked, {at, {at, John}}}} – *looked* projects

In this minimalist system, there is only one basic operation which combines elements and labels the merged object by making the head of the merged object project (Svenonius 2021: 143)[2].

As shown above in Table 1, the interpretation of a sentence can drastically change if we play around with the order, i.e., if we change the way words are merged. This can lead us to assume that elements must follow a strict ordering to preserve the intended meaning. However, this would be a false assumption. Language is a creative and complex system that allows diverse ordering of elements, while maintaining the intended interpretation. This point can be illustrated with the familiar example of the mouse being chased by the cat.

[2] More recently, it has been proposed that Merge is symmetric in that it creates unlabeled, unheaded sets. Instead a second operation *Label* is proposed to be active at the interface. Label labels syntactic objects for interpretability. This means that Merge is not responsible for labelling and headedness, but instead only combines lexical items into unordered sets (Svenonius 2021: 150).

Table 2: Word order varies, but the interpretation is the same.

String	*Interpretation*
a. The cat chased the mouse	
b. The mouse, the cat chased	
c. The cat who chased the mouse is…	
d. The mouse who the cat chased is…	
e. It was the mouse who the cat chased	
f. It was the cat that chased the mouse	

Table 2 illustrates that variability is permitted in how words are linearly ordered, while the basic interpretation of the sentence is preserved. In Table 2 (c and d) a relative clause is formed – the relative clause provides additional information about the cat (c) and the mouse (d). In Table 2 (e and f) an it-cleft is formed with *it was* followed by a relative clause. This construction focuses on the fact that it was *the cat* and not *the mouse* who did the chasing.[3] Thus, order of words not only yields differences in thematic roles, as in Table 1, but also interpretative differences that do not alter the interpretation of thematic roles as in Table 2. The difference between Table 1 (a and b) is the thematic roles assigned to the cat and the mouse. Thematic roles refer to the semantic roles played by each argument, e.g., who is instigating the action (agent) and who is undergoing the effect of the action (theme) or experiencing (experiences) som state etc. (Radford 2004: 250). In Table 1, we understand the cat as the chaser (agent) in (a), and the mouse as the chaser (agent) in (b).

Thus, language requires some rigidity in the way words are ordered, while at the same time allowing specific re-orderings. The word orders in Table 1 can be referred to as the canonical word orders for expressing the two interpretations of the cat chasing the mouse and the mouse chasing the cat. The possibility of displacement of phrases is a universal property of

[3] In the early years of GG under the EST of Chomsky (1973, 1977), sentences were thought to first be put together by specific rules, and then later transformed by other types of rules. This view provides the background for many of the terms and metaphors used to describe differences between sentences with the same interpretation. For instance, the term *topicalization* refers to the process of making an element a topic. This implies that the sentence is formed, and then one of the DPs for instance, is moved to a topic-position in the left periphery of the sentence. Many of these terms and metaphors are preserved in modern versions of GG. However, in MP there is no single operation *topicalization,* instead an element is made into a topic by the same basic operation, Merge, that combines all other elements.

language and one that has been one of the central research topics for a long time (Chomsky 2000: 23-26). Re-ordering operations are traditionally referred to as movement operations in the generative framework. With movement, the canonical word order is altered to change the interpretation, but the thematic roles are not changed. In MP these operations are reconceptualized as a type of Merge, meaning that there is only a single structure building operation, as opposed to several as in Government and Binding Theory (Adger 2021). I will continue to use the term "movement" in the current dissertation. This helps to distinguish this particular type of Merge (Internal Merge) from other applications of the operation (External Merge), but also helps to conceptualize *the effect* of this operation. These re-ordering operations, or movement operations, come in two different flavors. Operations that move elements to argument positions are referred to as A-movement, while operations that move elements to non-argument positions are referred to as A'-movement (read: A-bar-movement). The current thesis is concerned with the latter types of operations. (4) provides an overview of different types of A'-movement and an analysis of the underlying structures. The element that is displaced is provided in **bold** and the position in which it is interpreted is indicated by an underscore. The position of the underscore is typically referred to as *a gap* and the displaced element in bold, a *filler.*

(4)

Canonical word order:

John said to Mary that he hoped Liverpool would win the Champions League.

A'-Movement, non-canonical word orders:

a. Topicalization

   [CP **The Champions League**, [TP John said to Mary [CP that [TP he hoped [CP [TP Liverpool would win __]]]]].

b. Relativization

   [CP [TP They watched **the Champions League** [CP which [TP John said to Mary [CP that [TP he hoped Liverpool would win __]]]].

c. It-cleft

   [CP It was **the Champions League** [CP that [TP John said to Mary [CP that [TP he hoped [CP [TP Liverpool would win __]]]]]].

d. Question formation

[CP **What** [C did [TP John say to Mary [CP that [TP he hoped [CP [TP Liverpool would win __]]]]]]? [4]

An important prerequisite for being able to play around with these different A'-movements (i.e., ways of structuring information), is the ability to form dependencies between words at a distance (i.e., long-distance dependencies). In (4a), we understand *The Champions League* as the object of the verb *win* even though it does not appear in the "canonical object position" as it does in the base sentence – immediately following the verb. Within the generative framework (Chomsky 1965, 1986b) it is common to talk about these types of structures as though *The Champions League* has *moved* from the (now open) position immediately following the verb *win* (Ross, 1967). This implies that *The Champions League* was first generated in the position immediately following the verb before it moved to its pronounced position. In minimalist terms, we can say that *the Champions League* is merged in the object position, and that a copy of *the Champions League* is later merged in the topic position.

In more theory-neutral terms, we can say that there is a filler-gap dependency between the moved word and gap position. Looking at (4a) specifically, there is a dependency between the filler *the Champions League* and the gap in the embedded object position. Exactly how this is encoded differs between various formalisms.

It may appear as though A'-dependencies can apply freely to any structure. However, following early generative endeavors in which too many "possible human grammars" were made available (i.e., allowing structures that are not seen in any language), theorists sought to uncover constraints on possible grammars (Lasnik & Lohndal 2013: 28). Several types of constraints were uncovered, one of which, and arguably one of the most important empirical findings in theoretical linguistics (Boeckx 2008a: 151)[5], were constraints on A'-movement.

Ross (1967) was the first to describe sets of constraints on A'-movement in detail. He discovered that A'-movement is illicit if applied to certain domains, specifically complex NPs,

[4] In principle, any of the phrases in the base sentence can be A'-moved:

(i) Additional examples of A'-movement:

a. They watched **Liverpool** who John said that he hoped __ would win the Champions League.

b. **To whom** did John say __ that he hoped Liverpool would win the Champions League?

[5] "If asked what the most fundamental empirical discovery made by generative grammarians to this day is, I would unhesitatingly answer 'islands'" (Boeckx 2008a: 151).

coordinate structures and sentential subjects. Looking at the examples in (5), these are all instances of *rc*-dependency formation, which was shown to be licit in (4). Forming an *rc*-dependency into the following domains, however, is impossible:

(5)

a. Complex NP Constraint:
*__The hat__ which I believed [the claim that Otto was wearing __]
b. Coordinate Structure Constraint:
*Here's __the whisky__ which [I went to the store and Mike bought __]
c. Sentential Subject Constraint:
*__The teacher__ who [that the principal would fire __ was expected by the reporters] is a crusty old battleax.

(Ross 1967: 126, 168, 241).

To account for these facts, Ross (1967) formulates the Complex NP Constraint (CNPC), the Coordinate Structure Constraint (CSC), and the Sentential Subject Constraint (SSC). Both the CNPC and the CSC are proposed as universal constraints, while the SSC is language specific. As an example, we will consider the CNPC which reads as follows:

(6) CNPC:
"No element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation" (Ross 1967: 127).

The CNPC will rule out the sentence in (5a) because the noun phrase *the claim* dominates the finite clause *(that) Otto was wearing the hat*, and the CNPC prohibits any element inside a complex NP from moving out.

Ross (1967) gives the domains that do not allow A'-dependencies to be formed across their boundaries the metaphorical name "islands". The name alludes to the isolated status of these domains. In other words, in (5a) the complex NP is an island. Since Ross (1967), less categorical, but similar effects have been discovered in other domains. This has brought about a division of "island" into *strong* and *weak* islands. Strong islands are island domains that never allow extraction, while weak islands are thought to constitute domains that allow certain elements to escape. Incorporating weak islands into the catalogue of island domains has

massively expanded the number of domains that are considered islands (see Szabolsci & Lohndal 2017 for a thorough overview).[6]

1.1.1 The problem that islands pose

Although descriptively accurate, postulating this very specific constraint on A'-movement in a large set of different domains forces the question of how they come to be acquired by the speaker. The short answer has typically been that they are not acquired, but innate. This is generally assumed based on the logic of "the Poverty of the Stimulus" (POS): as there is no direct evidence available in the input that islands are unacceptable (i.e., island violations are not produced by speakers), children cannot reliably come to the conclusion that this configuration is unacceptable based on input alone, there must be some additional source of knowledge that contributes to their acquisition (Hornstein and Lightfoot, 1981: 12). Knowledge exceeding experience can be ascribed to the biologically endowed "learning mechanism" UG (Chomsky, 1986b: 18) (Lasnik and Lidz, 2017: 222). Within the generative framework, most theories rest on the assumption that the adult target state cannot be attained from inductive learning alone, and instead must be helped by UG (an idea popularized through Chomsky 1959, though as pointed out by Lohndal 2017, ch. 1: 1, originates in much earlier theories of language). Claims about universals of language have therefore been made in support of this theory (e.g., *that*-trace effects, Perlmutter 1968; Chomsky and Lasnik 1977), and islands have been a prominent example of such a universal constraint (Phillips 2013b: 132) as there does not seem to be any other way that the child can learn that these structures are not permitted. As Phillips (2013b: 132-133) points out island structures are obscure and abstract, and their effects are not easily identified in the input, which makes them difficult to learn from input.

Islands involve two phenomena that freely appear in natural language: (i) domains such as complex NPs, sentential subjects and coordinate structures, and (ii) long-distance dependencies such as top- and *wh*-dependencies. Both phenomena will appear in the child's input. However, there will not be positive evidence in the input that you *cannot* form long-distance dependencies into these domains. It is therefore a puzzle that speakers do not try to combine these phenomena. In addition, as pointed out by Sprouse & Hornstein (2013) and Phillips (2013b), there is (some) cross-linguistic uniformity regarding island constraints.

[6] I will not focus on this distinction as it implies that certain domains intrinsically belong to each of the categories. Foreshadowing somewhat, this dissertation argues instead that there is a distinction between derivational and representational islands, a distinction that includes an explanation of why certain domains allow some elements to escape and not others. All derivational islands should be domains that have been identified as strong islands, while representational islands will overlap with domains that have been identified as both strong and weak islands.

Phillips (2013b: 133) writes: "some island effects are sufficiently consistent across languages to be good candidates for universals; and those that do vary across languages appear to draw from a standard menu of options". As such, island effects seem to be good candidates for being part of UG as "the initial state of the language faculty" (Freidin & Lasnik 2011: 23).[7]

1.2 The focus of thesis: adjunct islands

Adjuncts have long been assumed to constitute classic examples of universal and strong islands. Huang (1982) was the first to point to a distinction in extraction patterns between non-complements (adjuncts and subjects) and complements, such that only complements allowed extraction. Stepanov (2007) and Truswell (2007, 2011) maintain that finite adjunct clauses, i.e., generally referring to finite adverbial clauses, are strong islands universally and should never allow extraction. This implies that the Adjunct Condition (AC – "the ban on extraction from adjoined phrases", Boeckx 2012: 16) might be a good candidate for an innate constraint. However, early research on Norwegian and Swedish indicated that not all finite adjunct clauses behave as islands (Anward 1982). This has recently been corroborated in research on Swedish (Müller 2019). The problem that this dissertation investigates is whether or not adjuncts as a class should be considered islands or not.

The dissertation has two main objectives – mapping the empirical landscape of adjunct islands in Norwegian; and investigating the origin of adjunct island effects, with the goal of providing a formal explanation of adjunct island effects. Adjunct islands have been investigated in the context of constraint-based approaches (i.e., the Subjacency Condition, the CED, *Barriers* (Chomsky 1986a)), yielding the hypothesis that all adjuncts are islands. However, the current state of the field suggests that there might be more variation in adjunct island effects than constraint-based approaches allow for. Current theories of adjunct island constraints must

[7] Under classic constraint-based accounts, island constraints are theorized to originate from innate constraints on language. The opposite assumption is to assume that island constraints are learnable from input (see discussion in Pearl & Sprouse 2013). If they are learnable from input, we assume that there must be some evidence for the constraints in the input. However, corpus searches only find evidence of a limited set of island violations (see e.g., Müllers & Eggers 2022). Unless we assume that learners can learn from negative input (i.e., the absence of a given construction), it must be assumed either that the constraints governing island effects are innate or that they are derived from more general constraints on language that are innate or potentially learnable from input. At the current state of the investigation, however, these hypotheses are not distinguishable. There is a growing field of computational linguistics in which these questions are investigated through learning algorithms (see e.g., Pearl & Sprouse 2013; Dickson, Pearl & Futrell 2022; Wilcox, Futrell Levy 2022; Kobzeva, Arehalli, Linzen & Kush 2023) which might have the potential to shed some light on this question.

be dramatically altered if variation in adjunct islands is found. Theories which have thus far provided a degree of explanatory power should only be altered in the face of rigorous empirical research providing evidence against theoretical predictions. Therefore, the starting point for the current thesis will be the following hypothesis:

(7) General hypothesis of the thesis
Extraction from adjunct clauses only minimally vary. In cases of variation, variation is systematic and can be explained by existing approaches to (adjunct) islands.

Several research questions arise from the hypothesis in (7).

(8) Research questions

Empirical research questions

a. Do different adjunct clause types behave in the same way with respect to long-distance A'-dependencies in Norwegian?
b. Do different long-distance A'-dependencies behave the same way with respect to different adjunct clause types in Norwegian?

Theoretical research question

a. How can the observed extraction patterns be analyzed formally?

The two empirical research questions are addressed in Papers 1 and 2. The theoretical research question is coarsely addressed in Papers 1 and 2, and investigated in detail in Paper 3.

1.3 Outline

The cover article is structured as follows. First, I provide a summary of the main questions and findings in Papers 1, 2 and 3. Here I show that there is fine-grained variation in extraction patterns in Norwegian (Papers 1 and 2) and provide a theoretical proposal to explain the variation (Paper 3). Next, Section 3.1 contains a detailed overview and discussion of different theoretical approaches to islands and adjunct islands specifically. Section 3.2 provides an overview of the empirical findings for adjunct islands cross-linguistically and discusses them in relation to the theoretical approaches presented in Sectio 3.1. In Section 4, I introduce the methodology used to collect data for Papers 1 and 2 and discuss potential challenges for the methodological design. Section 5 provides a detailed overview of the main findings of the

dissertation. Here I show the fine-grained patterns of variation uncovered in the empirical investigations, propose a theoretical explanation of the current findings in line with existing theoretical approaches to islands and explore the implications of the main findings. Section 5.4 discusses the current findings from a comparative perspective. Section 6 concludes the cover article and provides remarks on future work. Section 7 provides a broader view of how the three papers contribute to the field. Finally, the three papers follow.

2 Summary of papers

2.1 Paper 1

The first paper in the dissertation investigates extraction by topicalization out of three finite adjunct clauses in Norwegian. The paper is entitled "Variation in adjunct islands: the case of Norwegian" and is joint work with Dave Kush and Terje Lohndal. The paper follows up on a series of studies that report on finite adjunct island effects in different A'-dependencies (Sprouse, Wagers & Phillips 2012a; Sprouse, Caponigro, Greco & Cecchetto 2016; Kush, Lohndal & Sprouse 2018, 2019).

Finite adjunct clauses are often assumed to always be islands for filler–gap dependency formation cross-linguistically, but Kush et al. (2019) find experimental evidence suggesting that finite conditional *om* 'if'-adjunct clauses are not islands for topicalization in Norwegian. Paper 1 investigates the generality of Kush et al.'s (2019) findings by reporting and developing further statistical analyses on three formal acceptability experiments conducted in Bondevik (2018, MA thesis). We report on three acceptability judgment experiments testing topicalization out of three different adjunct clause types: conditional *om* 'if', habitual *når* 'when', and causal *fordi* 'because' in Norwegian to see whether the findings for conditional *om* 'if' revealed in Kush et al. (2019) generalize to different adjunct clause types.

First, the study largely replicates Kush et al.'s result (2019) for *om* 'if'. We find that topicalization from *om* 'if'-adjuncts yields small island effects in all three experiments. Second, the study provides evidence that the findings for adjunct *om* 'if' should not be generalized to other adjunct clause types. We find island effects of forming top-dependencies into *fordi* 'because' and *når* 'when'-adjuncts, but the size of the effects and the underlying judgment distributions differ between *fordi* 'because' and *når* 'when'. The results for *fordi* 'because' can be classified as classic island effects, while the average results for *når* 'when' disguise substantial underlying variation between trials. Subsequently, our results suggest that the syntactic category 'adjunct' may not constitute a suitably fine-grained unit to explain variation in island effects as we see differences in island effects between the three adjunct clause types.

2.2 Paper 2

The second paper in the thesis is joint work with Terje Lohndal and is entitled "Extraction from finite adjunct clauses: an investigation of relative clause dependencies in Norwegian". The paper follows up on the findings in Paper 1 that adjunct clause type matters for extraction. Moreover, it also follows up on findings that dependency type might matter for adjunct clause extraction (Sprouse et al. 2016; Kush et al. 2018; Kush et al. 2019). Sprouse et al. (2016) show

that adjuncts are not necessarily islands for all dependency types, and that conditional *if*-clauses do not yield classic island effects in English.

Paper 2 investigates whether the results in Paper 1 generalize to *rc* -dependencies in Norwegian by testing the same three adjunct clause types as Paper 1: conditional *om* 'if', habitual *når* 'when', and causal *fordi* 'because'. Moreover, as Paper 1 revealed small island effects for *om* 'if', Paper 2 investigates the origin of the small effect, specifically whether the small effect is caused by underlying variation.

We find that forming an *rc*-dependency into a finite adjunct in Norwegian overall yields 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' yields intermediate results. The study in Paper 2 largely replicates the study in Paper 1. In addition, we provide an answer for how the small island effect seen for *om* 'if' in both a top- and an *rc*-dependency can be interpreted. We find that the small island effect is not caused by variation, but that participants consistently rate extraction from conditional *om* 'if' as better than extraction from causal *fordi* 'because' and worse than extraction from declarative *at* 'that'-clauses. Thus, we argue that rather than relying on binary distinctions only, any theory that is to explain the empirical landscape must be sufficiently fine-grained allowing for more gradient distinctions between adjunct clause types.

2.3 Paper 3

The third paper in the thesis is entitled "Why adjuncts are not islands categorically: a case study of causal *fordi* 'because' and conditional *om* 'if' in Norwegian". This paper investigates and provides an explanation for the differences between conditional *om* 'if'-adjuncts and causal *fordi* 'because'-adjuncts focusing on the patterns revealed by the experiments in Papers 1 and 2. Papers 1 and 2 show that causal *fordi* 'because'-adjunct clauses yield classic island effects both in a top- and an *rc*-dependency. Conditional *om* 'if'-adjuncts, on the other hand, are shown to robustly yield small island effects in which the island-violating test sentence yielded intermediate ratings. In other words, where *fordi* 'because' does not allow extraction by way of topicalization or relativization, adjunct *om* 'if'-clauses are repeatedly shown to partially allow extraction, yielding acceptability ratings above *fordi* 'because', but below declarative *at* 'that'-clauses.

As of yet, no explanation has been offered that can explain the difference in extractability between adjunct clause types. The standard explanation for adjunct island effects is rooted in differences in external syntax between adjunct clauses and complement clauses, which builds on empirical work going back to Huang (1982). Paper 3 argues that the different

adjunct island effects of *om* 'if' and *fordi* 'because' can be derived without making reference to the external syntax of adjunct clauses. Instead, I argue that differences in internal syntax can account for the differences in extraction patterns. That the extraction patterns of two adjunct clause types, and the differences between them, can be explained by each adjunct clause's internal syntax is a novel proposal. I argue that this new perspective provides a promising ground for future work on adjunct islands as it allows us to explain patterns in adjunct islands without making additional assumptions about the nature of adjuncts. However, this proposal requires a theory of adjuncts as more syntactically integrated with the matrix clause than what is typically assumed (compare e.g., Chomsky 2000, 2001 on Pair-Merge).

Furthermore, the paper argues that two distinctive conditions on locality are required to arrive at a complete understanding of the extraction patterns. There is independent evidence that *fordi* 'because' occupies a position that blocks any phrase from moving out of the clause (i.e., the phase-edge, Chomsky 2000). This constitutes a derivational locality constraint. Next, there is evidence that conditional *om* 'if'-clauses have a different internal structure than *fordi* 'because'-clauses and that a *possible worlds operator* (Bhatt & Pancheva 2006) has moved to a specifier position in the left-edge to derive the conditional *om* 'if'-adjunct clause. In conditional *om* 'if'-clauses movement is possible from a structural perspective, but only so long as the moving element does not fully match the features of the possible worlds operator. This operator acts as an intervener for chain-formation (post-derivationally) for moving elements that fully match its features, and as a partial intervener for moving elements that partially match its features. Full overlap causes a large decrease in acceptability, while partial overlap only causes a slight decrease in acceptability (following Rizzi 2018). This constitutes a representational constraint. Accordingly, Paper 3 argues that we need both a derivational and a representational constraint on locality, and that in effect, *fordi* 'because' and *om* 'if' are not only different adjunct clause types, but different island types.

3 Background

This section provides an overview of the theoretical assumptions and empirical landscape of islands and adjunct islands in particular. In Section 3.1, I outline, review, and discuss the classic constraint-based explanations of islands that have been proposed in succession starting with the Subjacency Condition. I focus primarily on explanations of adjunct islands. In Section 3.2, I provide an overview of the empirical status of adjunct islands cross-linguistically, with a particular focus on Norwegian and Mainland Scandinavian languages. Next, I argue that the theoretical predictions of the classic constraint-based approaches do not overlap with the empirical landscape. I briefly review more flexible approaches that have been proposed to account for unexpected variation in adjunct island effects.

3.1 Approaches to islands

There are two main views of why islands arise: the traditional grammatical view and the processing-based view. On the traditional grammatical view, islands constitute constraints on the grammar, either syntactic, semantic, or pragmatic. In other words, islands are domain-specific (Phillips 2012a: 64). This implies that the constraints must somehow be part of grammar. In the processing-based view, which is also referred to as the reductionist approach, islands are thought to arise as an effect of an overloaded processing device. This places islands outside of grammar and in a domain-general cognitive area. As such, the grammar is simplified. There is ongoing debate as to which of these views is correct, and importantly, the views are not mutually exclusive. It is possible that certain island effects which have been identified might be constrained by limitations on processing, while others are constrained by the grammar. I return to this question in Section 4. For now, as islands were first discovered within a grammatical approach, I will focus on the traditional grammatical view. In a grammatical approach, based in the generative framework, islands are thought to be derived from linguistic universals that are typically assumed to be syntactic in nature (Sprouse et al., 2012a). The job of the linguist is to identify these linguistic universals. Building on the generalizations and descriptions provided by Ross (1967), linguists within the generative framework seek to derive the conditions on A'-movement from general structure building mechanisms. Since Ross (1967), there have been three principal implementations – the Subjacency Condition, *Barriers* (Chomsky 1986a) and Phase Theory. These can collectively be referred to as *constraint-based approaches to locality* and they all assume that (i) movement is successive cyclic, and (ii)

islands arise if any obstacle (filled COMP-node, barrier, filled phase-edge) inhibits movement from occurring in small steps.[8]

3.1.1 Subjacency

Chomsky (1973) continues Ross' (1967) work on islands and unifies the different domain specific constraints in one general and universal constraint: the Subjacency Condition.[9] The Subjacency Condition is defined as a general constraint on transformational operations, stating that there must be a specific relation between the constituent that moves and the landing site in order for the move to be licit: the moving constituent and the position to which it moves must be *subjacent*. In (9) below, Y can be understood as the mover and X as the landing site:

(9) The Subjacency Condition:

> No rule can involve X, Y, X superior to Y, if Y is not subjacent to X.
>
> (Chomsky 1973: 247)

In terms of movement operations, *to be subjacent* means that there are less than two cyclic nodes between the filler and the gap. In Chomsky (1973), S' and NP are taken to be cyclic nodes (corresponding to CP and DP in modern frameworks):

(10) To be subjacent

> No rule can move an item from position Y to position X in the structure
>
> … [β … [α … Y …] …] … X …
>
> Where Y is not α and α, β are cyclic categories, unless some constant terms of the structural description of the rule holds of a phrase in β that is subjacent to X.
>
> (Chomsky 1973: 271)

[8] Boeckx (2012: 58) sums up the constraint-based approaches excellently in the following way: "[…] the syntactic component is designed in such a way that operations must take place within a narrow computational window (amounting to the size of the syntactic cycle); in particular, only nice short movements are tolerated. This is what gives rise to the phenomenon of successive cyclic, step-by-step movement. Island effects arise when movement is forced (for one reason or another) to be longer than it should, transgressing the limits imposed by the cycle. This is the reigning idea behind the notion of bounding node, barrier, and, more recently, phase".

[9] Chomsky (1973) also proposes the Tensed S Condition and the Specified Subject Condition to supplement the Subjacency Condition. The former constrains operations that cross a tensed clause, and the latter ones that cross a specified (i.e, overt, lexical – Boeckx 2012: 9) subject. These will not be discussed any further in this cover article.

Simply put, no operation can relate the position of Y, which is embedded below two cyclic nodes with the position of X, which is outside of the domain of the same two cyclic nodes. In other words, any operation that crosses two cyclic nodes is prohibited by the Subjacency Condition. In (11), it would be impossible to move Y to X given Chomsky's (1973) postulation that S' and NP are cyclic nodes.

(11) […X …[S' … [NP … Y …]]]

There are, however, countless examples that show that Y can be moved to X even across more than two cyclic nodes. The example in (12) shows that *what* can cross three cyclic nodes (S') and the sentence is still acceptable.

(12) **What** [S' did John remember [S' that Peter had said [S' that he wanted to buy _]]]?

Though this sentences seemingly is a violation of the Subjacency Condition, it is still acceptable. It is assumed that *what* stops over in the intermediate position of the left-edge of each clause, i.e., *what* moves successive cyclically. Successive cyclicity means that there are consecutive cycles of movement in a derivation. Thus, instead of postulating that *what* moves directly outside of its clause in (12), it is typically assumed is that *what* first moves to the closest S' (i.e., the first cyclic domain) before moving again. In other words, instead of undergoing *one* long movement, *what* undergoes several shorter moves. At first glance this might appear to be an ad hoc solution to the problem posed by (12). However, there are several pieces of empirical evidence from different languages that provide proof of concept for successive cyclicity, perhaps most notably Belfast and West Ulster English (Henry 1995 and McCloskey 2000, repspectively) (see also den Dikken 2009 on restructuring effects). Thus, *what* in (12) is postulated to only cross one S'-node at a time, as illustrated in (13). In Chomsky (1973, 1977) the intermediate position is implemented as the left-most node in the clause (COMP in S' in Chomsky 1973, 1977).

(13) **What**$_i$ [S' $_i$ did John remember [S' $_i$ that Peter had said [S' $_i$ that he wanted to buy $_i$]]]?

If the A'-moving element makes it to a COMP-node in one cycle (i.e., by one application of A'-movement), then this element is allowed to move on to the next cycle. The Subjacency

Condition will correctly predict that (13) is grammatical as *what* only crosses one S' each time it moves.

The Subjacency Condition makes the correct predictions for grammatical sentences, but can it also explain why some sentences are ungrammatical? In particular, can it constrain movement from islands? Revisiting the CNPC, we see that the Subjacency Condition can explain why extraction from such a structure is illicit, while extraction from a very similar structure is licit.

(14)

a. *[**The hat**]$_i$ [$_{S'}$ which I believed [$_{NP}$ the claim [$_{S'}$ _$_i$ that Otto was wearing _$_i$]]]

b. [**The hat**]$_i$ [$_{S'}$ which I believed [$_{S'}$ _$_i$ that John claimed [$_{S'}$ _$_i$ that Otto was wearing _$_i$]]]

In (14a), *the hat* moves to the left most position of the first clause. On the next cycle, as the left most position below the matrix S' is occupied by *which*, *the hat* must cross both the NP and the S'. This is constrained by the Subjacency Condition. In (14b), *the hat* moves successive cyclically to the left-most position of each clause, with exception of the last clause, as it is occupied by *which*. However, there is only one cyclic node (S') that *the hat* must cross in one cycle, and subsequently, this move is not constrained by the Subjacency Condition.[10]

Interestingly, subjacency, which is a general constraint on language, can constrain the movement in (14) without having to make claims about complex NPs specifically. This is an advantage of the Subjacency Condition compared to Ross (1967). Chomsky (1977: 118) discusses whether the CNPC can be abandoned in favor of the more general Subjacency Condition and finds that it can and should as there are cases of permissable extraction that are ruled out by the CNPC, but allowed by the Subjacency Condition.

[10] Chomsky (1973/1977) does not provide any independent motivation for why some nodes are considered to be cyclic and others not. In fact, he does not provide any definition of a 'cyclic node'. It is only through investigating how well Subjacency provides the right empirical outcomes for transformational operations that he builds evidence that the choice of cyclic nodes is correct. Chomsky (1964: 39) makes it clear that to choose between different hypotheses is to ask which hypothesis of language provides descriptively adequate grammars for natural language. Chomsky (1977) revises the hypothesis that S' and NP are cyclic nodes. Based on additional data, he shows that postulating that S and NP (TP and DP in modern frameworks) are cyclic nodes provides the right outcome (i.e., a descriptively adequate grammar).

3.1.2 *Barriers*

As the Generative Framework evolved from the transformational grammar of Chomsky (1973, 1977), the Subjacency Condition was also replaced. Under Government and Binding (GB) (Chomsky 1981) the cyclic nodes of the Subjacency Condition are reimplemented as barriers to movement (Chomsky 1986a). The idea inherent to the Subjacency Condition, however, is maintained – movement from one position to another cannot happen across too great a distance. One of the driving forces behind the reformulation of cyclic nodes to barriers is the discovery of important constraints on movement that the Subjacency Condition, as defined in Chomsky (1973, 1977), cannot account for. The main insight is presented in Huang (1982). He shows that there is a distinction between movement from a complement position and from a non-complement position, which he unifies in the Condition on Extraction Domain (CED):

(15) Condition on Extraction Domain (CED)
A phrase A may be extracted out of a domain B only if B is properly governed.
(Huang 1982: 505)

Without going into details of what (proper) government is, an empirical generalization is that complements are properly governed domains while and adjuncts and subjects are not properly governed domains.[11]

(16) Difference between complement extraction and non-complement extraction
a. **Which book**$_i$ did you buy $\__i$ without reading $\__i$?[12] (Complement)
b. ***Which book**$_i$ did you go to college without reading $\__i$? (non-complement: adjunct)
(Huang 1982: 505)

There is no structural difference between complements and non-complements with regards to the configurations that Subjacency under EST is concerned with, i.e., the sequence of nodes that head these constituents. This can be illustrated with the machinery available in the GB framework: the difference between (16a) and (16b) is that the path to the matrix Spec-CP

[11] Boeckx (2012: 21) writes that (proper) government is "[…] a structural relation that received many modifications during the 1980s, due to the large amount of data that it was meant to cover, and that syntacticicians have been trying to avoid ever since". I follow Boeckx (2012) and will generalize the CED to the distinction between complement and non-complement.

[12] The lowest gap is here a parasitic gap (Engdahl 1983), which Huang treats as being base generated in this position. Thus, there is no related element that has been extracted from this position, i.e., no violation of the CED.

position from the gap position is CP-IP-VP in the former and CP-IP-VP-CP-IP-VP in the latter. However, the diverging paths cannot explain the difference in grammaticality. The path in (17) below is also CP-IP-VP-CP-IP-VP, however, unlike (16b) the dependency between the filler and the gap is successfully formed:

(17) [CP **Which book**$_i$ did [IP John [VP think [CP that [IP Mary [VP had bought $_i$]]]]]]?

Looking at these examples, it is clear that the Subjacency Condition under EST is unable to capture these differences as it will predict that (16a, b and 17) all will be cases of acceptable movement. In (16b), the *wh*-word only crosses one cyclic node in each cycle, as the most embedded Spec-CP is open. The same occurs in (17). In other words, the examples in (16b) and (17) both involve movement across exactly one cyclic node.

Barriers, however, introduces distinctions in the definition of the barrier nodes that can explain the difference in acceptability. The real difference between the cyclic (bounding) nodes and the new barriers, is that there are no nodes that inherently are barriers to movement, like S and NP in Chomsky (1977). Nodes will only be barriers to movement in specific environments and not in others. Huang (1982) proposes that the difference between an island domain and a non-island domain lies in how each relate to the other nodes in the sentence – complements enjoy a particularly close relationship with the matrix clause, while non-complements are more peripheral. It is precisely this insight that is captured by *Barriers,* as the definition of a barrier is relational:

(18) Definition of a barrier

> "a category α will be a barrier for β for certain choices of β but not for others; a category may be a barrier by inheritance or intrinsically" (Chomsky 1986a: 12).

A barrier is defined in GB-specific terms such as L-marking and Blocking Categories (BC). What follows are definitions of L-marking, BCs and barriers:

(19) L-marking

α L-marks β iff α is a lexical category that θ-governs[13] β.

(Chomsky 1986a: 15)

(20) Blocking category

γ is a BC for β iff γ is not L-marked and γ dominates β.

(Chomsky 1986a: 14)

(21) Barrier

γ is a barrier for β iff (a) or (b)

a. γ immediately dominates δ, δ a BC for β;

b. γ is a BC for β, γ is not IP

(Chomsky 1986a: 14)

(21a) dictates that γ inherits barrierhood from a BC that it dominates, while (21b) dictates that γ is a barrier intrinsically by virtue of being a BC. These are quite intricate and abstract definitions that will benefit from illustrative examples. In (22a), the head of the adjunct clause is a BC by (20) because it is not L-marked, thus, it is a barrier by (21b). Furthermore, IP inherits barrierhood by (21a) as it immediately dominates VP which is a BC by (20). Consequently, for *who* to move to the Spec-CP position of the matrix clause, it must violate subjacency by crossing two barrier nodes. Similarly, in (22b) the head of the relative clause will have to move across two barriers in one move: the embedded subject is a BC by (20) as it is not L-marked, thus it is also a barrier (by 21b). Next, IP inherits barrierhood by (21a).

(22) Examples of how subjacency will work with relational barrierhood in *Barriers*

a. ***Who** did [$_{IP}$ they leave [$_{CP}$ before speaking to _]]?

b. ***The man** who [$_{IP}$ [$_{NP}$ pictures of _] are on the table]

(Chomsky 1986a: 31)

The examples in (22) are exactly of the type that Huang (1982) discusses and which the EST Subjacency Condition is unable to explain: extraction from non-complements, an adjunct clause in (22a) and a complex subject in (22b). Subjacency under *Barriers*, however, is able to rule

[13] θ-marking is defined in Chomsky (1986a: 15) as follows:

"α theta-governs β iff α is a zero-level category that θ-marks β, and α, β are sisters".

out these sentences. As such, Chomsky (1986a: 34) argues that the CED can be reduced to subjacency.

Yet, there is a problem. In (23) below, both sentences are grammatical. In (23a), the adjunct clause only crosses one barrier IP, which should be fine under subjacency. However, the complement in (23b) crosses two barriers on its way to Spec-CP; VP is a BC and a barrier and IP is a barrier by inheritance.[14] Such movement is incorrectly ruled out by subjacency.

(23) Apparent problem for *Barriers*

a. **How** did [$_{IP}$ you [$_{VP}$ fix the car] _]

b. **Who** did [$_{IP}$ John [$_{VP}$ see _]] (Chomsky 1986a: 28)

Chomsky (1986a: 29) solves this by postulating that an A'-moving element can successive cyclically adjoin to VP. As such, the A'-moving element will only cross one barrier at a time as can be seen in (24) below. This possibility is not available in (22a) above, as the adjunct is immediately dominated by IP and not VP.

(24) Adjunction to VP

Who$_i$ did [$_{IP}$ John [$_{VP}$ _$_i$ [$_{VP}$ see _$_i$]]]? (Chomsky 1986a: 29)

Allowing adjunction to VP means that there are in principle two stopping points in the derivation: VP and CP. Part of the progression from the initial stages of EST was to reduce the various transformational operations to only constitute "move α". Move α, basically means "take anything and put it anywhere else, leaving behind a trace" (Lasnik & Uriagereka 1988: 5). Thus, all types of movement are unified under this general transformation. Accordingly, several instances of move α must be allowed within one cycle, but each constituent is only allowed to move once per cycle. Chomsky (1986a) posits that a *wh*-word can move to two different positions: adjunction to VP if that is in reach and likewise movement to Spec-CP. Thus, he assumes that VP and CP each constitute a cyclic domain.

Additionally, relevant for Paper 2, in particular, Chomsky (1986a: 28) postulates that there is some gradience in long-distance dependency formation such that crossing two barriers will cause the most decrease in acceptability, one – slight decrease and zero – no decrease: "[…] movement should become "worse" as more barriers are crossed, the best case being the crossing

[14] (17) is also incorrectly ruled out by subjacency under *Barriers* in the same way.

of zero barriers". As no adjunct clause is properly governed, the mover trying to escape an adjunct must cross two barriers on its way to the matrix Spec-CP – the adjunct clause will always be a barrier and the node to which it is adjoined will be a barrier by inheritance (e.g., IP in (25)). Thus, by default, extraction from an adjoined clause will yield a large decrease in acceptability. This is (a simplified version of) how subjacency under *Barriers* rules out movement in (16b) above, repeated here as (25):

(25) *[CP**Which book** did [IP:BARRIER by inheritance: you [VP go to college] [CP:BARRIER: without reading _]]]]?

3.1.3 Phase Theory

Phase Theory in the Minimalist Program constitutes a second reformulation of the original ideas of the Subjacency Condition. This "new" system, as pointed out by several authors, very closely resembles the system put forth in *Barriers* (Chomsky 1986a) and EST (Chomsky 1973, 1977) (Boeckx & Grohmann 2007, Boeckx 2008b, Gallego 2012). A *phase* roughly corresponds to a barrier in *Barriers*. However, the phase domains are more strictly determined than in *Barriers,* echoing the rigidity of cyclic nodes (Boeckx & Grohmann 2007: 218). The basic idea behind *phases* is that syntactic derivations proceed in incremental chunks, i.e., *phases*, that each are built from separate lexical sub-arrays (Chomsky 2000: 106, Boeckx & Grohmann 2007: 205). This means that each phase is built separately, and that the lexicon, which is thought to constitute a heavy load on the workspace (Chomsky, 2000: 100-101), is only active at certain points in the derivation (Gallego, 2012: 11). In other words, building structures in phases is assumed to alleviate the burden on the computational system (Boeckx & Grohmann, 2007: 206).

The original proposal is that *v* and C are phase heads (Chomsky 2000: 106). Chomsky's motivation for postulating *v* and C as phase heads is that these nodes are the closest syntactic counterparts to propositions – *v*P is a verb phrase in which all theta-roles are assigned, and CP is a full clause including tense and force (2000: 106). Motivation for postulating these to be phase heads is empirical evidence that phase heads trigger movement to their respective edges. Thus, phasehood is also established by "stop-over"-effects. "Stop-over"-effects are effects triggered by successive cyclic movement of an element to an interim position e.g., reconstruction effects, remnant quantifiers etc. In the examples in (26) and (27) given below, it is illustrated that in West Ulster English, it is acceptable to move *what all* together to the matrix CP (26) or to pronounce a remnant or associate of the filler at a lower point in the movement

path (27). This is taken to indicate that the *wh*-word was once placed in this lower position in the path, before it moved on to the next cycle, leaving behind a remnant.

(26) Pronunciation of a remnant of the filler and the filler in a high position

[CP [DP **what all**]$_i$ did [IP he say [CP _$_i$ that [IP he wanted _$_i$]]]]

(West Ulster English, McCloskey 2000: 61)

(27) Pronunciation of a remnant of the filler in a low position

[CP **what**$_j$ did [IP he say [CP **[DP _$_j$ all]**$_i$ that [IP he wanted _$_i$]]]]

(West Ulster English, McCloskey 2000: 61)

There are several pieces of empirical evidence from different languages that provide evidence for successive cyclicity; however, that *v*P and CP specifically constitute default phase edges cross-linguistically is a stipulation made by Chomsky (1995) and is not necessarily substantiated by the evidence for successive cyclicity available.[15] There is considerably more evidence that CP is a phase-edge than evidence that *v*P is a phase-edge. For my purposes, CP will be the important phase-edge and therefore, in this dissertation, I will not take the phasehood of *v*P into account.

Simplified for current purposes, movement within *Phase Theory* will proceed as follows: each phase is built from a lexical sub-array. The words that need to move out of the phase must make their way to the phase edge during the course of the derivation of the phase, similar to movement to Spec-CP and COMP-position in *Barriers* and EST, respectively. This is because the phase edge and the phase head are the only positions that are postulated to be visible for the later phases. This is known as the *Phase Impenetrability Condition* (PIC) (Chomsky 2000: 108).

(28) *Phase-Impenetrability Condition* (PIC):

In a phase α with head H, the domain of H is not accessible to operations outside α, only H and its edge are accessible to such operations. (Chomsky 2000: 108)

[15] We only have evidence of punctuated paths, but we do not necessarily have strong evidence for which nodes constitute the intermediate points in a derivation (Mike Putnam, p.c.). Though this indicates that we might need a more fluid definition of punctuated paths than what Phase Theory allows, I will continue to use Phase Theory in the current thesis as this is currently the most developed minimalist approach to cyclicity.

The PIC enforces strict cyclicity and yields a strong form of Subjacency (Chomsky 2000: 108). This can be illustrated in (29) below. *what* is base-generated in VP. As is evident from (29 d and j), each time CP (phase-edge) is merged in the structure, the *wh*-word moves to the Spec-CP. In (29j) *what* moves to the matrix phase edge and the derivation is complete.

(26) Derivation of movement in Minimalism's *Phases*

a. [$_{vP}$ John [$_{v'}$ *v* [$_{VP}$ eat what]]]
b. [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat what]]]]
c. [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat what]]]]]
d. [$_{CP}$ **what**$_i$ [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat _$_i$]]]]]]
e. [$_{VP}$ say [$_{CP}$ **what**$_i$ [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat _$_i$]]]]]]]
f. [$_{v'}$ *v* [$_{VP}$ say [$_{CP}$ **what**$_i$ [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat _$_i$]]]]]]]]
g. [$_{vP}$ you [$_{v'}$ *v* [$_{VP}$ say [$_{CP}$ **what**$_i$ [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat _$_i$]]]]]]]]]
h. [$_{TP}$ **you**$_j$ [$_{vP}$ _$_j$ [$_{v'}$ *v* [$_{VP}$ say [$_{CP}$ **what**$_i$ [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat _$_i$]]]]]]]]]]
i. [$_{C'}$ did [$_{TP}$ **you**$_j$ [$_{vP}$ _$_j$ [$_{v'}$ *v* [$_{VP}$ say [$_{CP}$ **what**$_i$ [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat _$_i$]]]]]]]]]]]
j. [$_{CP}$ **what**$_i$ [$_{C'}$ did [$_{TP}$ **you**$_j$ [$_{vP}$ _$_j$ [$_{v'}$ *v* [$_{VP}$ say [$_{CP}$ _$_i$ [$_{C'}$ that [$_{TP}$ **John**$_k$ [$_{vP}$ _$_k$ [$_{v'}$ *v* [$_{VP}$ eat _$_i$]]]]]]]]]]] (based on Hornstein et al. 2005: 360-361)

Thus, as illustrated in (29), the derivation of long-distance dependencies within MP will, by nature of being built in phases, be cyclical.[16] The derivation in (29) looks very similar to what has been postulated in previous theories (see discussion in Boeckx & Grohmann 2007 and Boeckx 2008b).

The PIC dictates that the intermediate landing site must be in the left edge of each of the phase heads. The consequence of the PIC is that no domain in which Spec-CP is open, will be an island. That being so, any domain where there is a filled Spec-CP will be an island. In (30) below, we see that the embedded Spec-CP is filled by *which earrings*. Thus, *who* cannot move out of the embedded *wh*-clause.

[16] Within the MP, most agree that movement is cyclic, and specifically, that it is successive cyclic. However, there is still a lively debate as to the exact details, such as the movement path from gap to landing site and triggers for movement. I will not review this debate as it is not directly relevant for the current project (but see among others Chomsky 2000; Gallego 2012; Abels & Bentzen 2009; Boeckx 2008b; Boeckx & Grohmann 2007; den Dikken 2009; Bošković 2007; Lasnik 2003, 2006; Richards 2011 for interesting discussions of these controversies).

(30)

a. [CP [TP Magnus [VP wonders [CP **which earrings**$_i$ [TP Ingrid would [VP like $\_i$]]]]]]]

b. [CP ***Who**$_j$ [C' does [TP Magnus [VP wonder [CP **which earrings**$_i$ [TP $\_j$ would [VP like $\_i$]]]]]]]?

Unlike in *Barriers* (Chomsky 1986a), no gradience is postulated in *Phases*. An element either can or cannot move to the phase edge – if it can the domain is not an island, if it cannot, the domain is an island.

A problem for Phase Theory, however, as for the Subjacency Condition, is the evidence that adjunct clauses seemingly are islands. Phase Theory cannot explain this data without additional assumptions about the phase status of adjuncts. I will review two main types of approaches to the Adjunct Condition (AC) in the MP. Both take as their point of departure the nature of adjuncts and particularly how they differ from complements. The problem for minimalist accounts is that as few operations as possible should be introduced in the theory and this makes it difficult to account for adjuncts which have a set of adjunct-specific properties that no other constituent group shares. Both approaches make use of Merge to account for the AC by assuming separate properties of Merge for adjuncts[17]: (1) the late-Merge approach (Stepanov 2001) and; (2) the invisible merge approach (Pair-Merge; Chomsky 2000, 2004). Both of these can be labelled as "separate-dimensions"-approaches following Bode (2020).[18] These accounts, building on Huang's (1982) original insight, see adjuncts as less integrated compared to complements, belonging in a separate space from the rest of the clause, even after Merge has applied. Going through these approaches in more detail, I will focus on how they explain the original assumptions of the AC namely that all adjuncts are islands.

The late-Merge approach

In the late-Merge approach, the timing of Merge is manipulated to distinguish between adjuncts and complements. The assumption is that adjuncts are merged later in the derivation than

[17] Adger (2003) assumes that there is another basic operation called *adjoin.* Adjoin incorporates adjuncts into the sentence such that YP adjoin to XP means that XP is copied and YP is inserted between the two XPs. This, however, introduces a second type of basic operation to the system, and could be argued not to be true to minimalist desiderata.

[18] There are other approaches to adjuncts, for instance analyzing adjuncts as specifiers of functional heads (cartography – see Cinque 1999) or analyzing adjuncts as predicates that select main clauses (the predication hypothesis, see e.g., den Dikken 2006). These types of approaches will not be discussed in this cover article as they have not been applied to adjuncts and islands to any extent.

arguments (Lebeaux 1988). This implies that merging of adjuncts is a post-cyclic operation (Stepanov 2001: 94). The late-Merge hypothesis finds support in evidence that adjuncts do not reconstruct, whereas complements do[19]:

(31) Evidence that adjuncts do not show reconstruction effects with Binding Condition C:

a. ?*Which argument that John$_i$ is a genius did he$_i$ believe?

b. Which argument that John$_i$ made did he$_i$ believe?

(Stepanov 2001: 95)

Binding Condition C states that R-expressions must be unbound in their domains (Chomsky 1981). In (31a), *John* cannot be interpreted as *he* in the matrix clause because *John* is within the domain of *he*. The reconstruction effects indicate that the original position for *which argument that John is a genius,* where *John* is c-commanded by *he* is available at the time when the NP is moved. In (31b), however, *John* inside the adjunct clause and *he* in the matrix clause can refer to *John*. Binding Condition C is assumed to apply at all stages of the cycle and thus (31b) should be a violation of Binding Condition C. This is not the case, however, and accordingly, such data have been taken as evidence that adjuncts are merged at a late stage, i.e., post-cyclically (Stepanov 2001).[20]

If we assume Late-Merge, then extraction from adjuncts should not be possible as post-cyclic movement is generally not assumed to be permitted (though see Hornstein & Nunes 2008). Given that adjunct clauses typically are considered as the prototypical case of a strong island, this is a desirable result. We can now explain the AC in Minimalist terms:

---

[19] Reconstruction here means that an interpretation that requires a local binding relation is available even after a given element has moved away from this position, i.e., the surface structure does not give rise to the interpretation. However, the fact that the interpretation is available indicates that certain elements inhabited different positions at an earlier stage in the derivation, making other local relationships possible.

[20] Bode (2020) and Stepanov (2001) show that the examples in (31) might be a simplification. In (ii) we see evidence that there might be reconstruction effects in the adjunct and not in the complement, in direct opposition to the assumption that adjuncts are merged at a later stage.

(ii) Counterexamples to reconstruction effects

a. *[Go home [after John$_i$ arrived]] he$_i$ did

b. [These pictures [of John$_i$]] seem to him$_i$ to be on sale

(Bode 2020: 14-15):

(32) *Who did Mary cry after John hit?

Step 1 – cyclic: [CP [C **did** [TP Mary <*did*> [vP [VP cry]]]]]

Step 2 – cyclic: [CP **who** [C after [TP John [vP [VP hit <*who*>]]]]

Step 3 – post-cyclic: [CP [C **did** [TP Mary <*did*> [vP [vP [VP cry]] [CP **who** [C after [TP John [vP [VP hit <*who*>]]]]

The reason that *who* cannot move from the adjunct clause *after John hit who* is that once the adjunct is merged with the matrix clause, cyclic operations such as A'-movement are completed, and the derivation stops without *who* being able to reach the matrix Spec-CP (see Stepanov 2007). This is an example of an approach that manipulates the temporal dimension of adjunct merger. Below, I review an approach which distinguishes spatial dimensions of adjuncts and complements.

The Pair-Merge approach

Chomsky (2000, 2004) postulates that merge of adjuncts, as merge of complements, is cyclic. Thus, there must be some other distinction between the merge of adjuncts and complements. Chomsky (2000, 2004) draws a distinction between a merged object that is ordered and one that is unordered, calling the former Pair-Merge and the latter Set-Merge. The difference between adjuncts and complements is here defined as the difference between the ordered pair <α, β> and the unordered set {α, β} (Chomsky 2000: 133). Under Pair-Merge, merging α, β yields {γ, <α, β>}, where γ is the label for β if α is adjoined to β. Thus, the adjunct (α) is invisible for further derivation as the object it is Pair-Merged with (β) projects. Under Set-Merge there is an unordered pair, both of which can project. Bode (2020: 53) sums up this distinction as the difference between (33a and b):

(33) Different types of Merge

a. Asymmetric Set-Merge: {X, Y} = {Y, X} => unordered pair
b. Symmetric Pair-Merge: <X, Y> ≠ <Y, X> => ordered pair

A further distinction between these two types of Merge is that Set-Merge is obligatory and the selector projects its label, while Pair-Merge is optional and has no selector.

(34) {hit, {hit, John}}
<{hit, {hit, John}}, hard>
{hit, <{hit, {hit, John}}, hard>}

In (34), *hit* is projected because it selects *John.* Then the adjunct *hard* is Pair-Merged with the object {hit, {hit, John}}. The host of adjunction, here *hit John*, retains all its properties. Thus, a VP that hosts an adjunct AdvP remains a VP. The adjunct AdvP behaves "as though it is invisible" (Bode 2020: 53). This situation is described by Chomsky (2004: 117-118) as the adjunct α being attached to β "on a separate plane, with β retaining all its properties on the "primary plane," the simple structure". If we come back to the familiar example repeated here in (35), we can now explain why extraction is impossible:

(35)

[CP [C **did** [TP Mary <*did*> [*v*P [*v*P [VP cry]]] ~~[CP **who** [C after [TP John [*v*P[VP hit <*who*>]]]…~~

visible structure | *invisible structure after Pair-Merge*

As the adjunct is Pair-Merged with the *v*P, only the *v*P projects and is visible for further operations. The adjunct is present at the time of cyclic operations, but on a dimension that is invisible for the rest of the derivation and subsequently it cannot be part of any cyclic operations. In fact, Chomsky (2000, fn. 102) assumes that adjunction is not part of narrow syntax (i.e., in the derivational space). Later, as part of Transfer, an operation SIMPL(ify) converts the ordered set into a standard set which means that elements from the separate plane are incorporated into the simple structure for interpretation at the interfaces (Bode 2020: 56). There are several unanswered questions about how this works. For one, it is not clear how the system "decides" between Set- and Pair-Merge. One suggestion is that there is a silent head (Mod) on each adjunct clause which is a functional layer subjected to Pair-Merge (Rubin 2003) yielding <{Mod, PP}, NP> as the ordered pair and {{Mod, PP}, NP} after Simpl has applied. Bode (2020: 55) criticizes this solution. She writes that the Mod head would have to select an array of different categories, i.e., any category which can be an adjunct, which is highly unusual for functional heads. The introduction of Pair-Merge into the theory is made both on empirical and minimalist grounds. Chomsky argues that a distinction between arguments and adjuncts is an empirically necessary distinction, which means that there must be a way within the system to distinguish the two. Pair-Merge evades the problem of having post-cyclic Merge, which can

be argued to contradict minimalist assumptions. Still, Pair-Merge forces us to postulate two types of Merge, which as Bode (2020: 54) points out is a complication of the theory.

3.2 Empirical data

3.2.1 Empirical predictions

All of the approaches reviewed above provide quite clear-cut empirical predictions: adjuncts will always be islands, regardless of the type of adjunct, type of language or type of A'-dependency. This is mainly due to two characteristics of these approaches: (i) the AC follows from assumptions that adjuncts as a phrasal category are distinct from other phrasal categories. The basic facts that make adjuncts different from complements, e.g., that adjuncts are "less integrated" in the main structure, are deployed to capture the empirical findings that adjuncts are strong islands. This means that it is the "adjunct"-property of these clauses that is relevant for these approaches, and not the semantic interpretation of the adjunct, the place of merge of the adjuncts etc. (ii) The approaches do not make fine-grained distinctions between type of adjunct, type of language or type of A'-dependency. As such, the assumption is that such fine-grained variation is not relevant for islandhood. Accordingly, given these accounts, we predict that all adjuncts will be islands categorically.

The null hypothesis established by these analyses is that there will be a binary difference in acceptability of extraction between adjuncts and complements as (i) there is no gradience in the notion of either being an adjunct ("island") or a complement ("non-island") (CED); (ii) in the notion of being properly governed ("non-island") or not ("island") (Chomsky's *Barriers* 1986a); (iii) an element can either be in a separate dimension (either temporally or spatially) or be in the same dimension – in other words, there is no middle ground. Thus, all else being equal, we predict that there will be big differences between the acceptability of extraction out of an adjunct and a complement. In addition, for extraction, it is predicted that all adjuncts should yield the same acceptability scores on one end of the scale, and all complements on the other end.

Another empirical prediction is that adjuncts will be the same cross-linguistically. The POS logic assumes that island constraints are innate and cannot be learned. Thus, unless there are independent reasons to postulate that there are inter-language differences on the general property of adjuncts, which are the mechanisms by which the AC is derived in the MP, cross-linguistic variation is not predicted. If we do, however, find that there is cross-linguistic variation that cannot be explained by language internal properties, we must assume that the AC somehow is acquired by the speaker after all. Thus, postulating cross-linguistic unity is the most

minimal hypothesis, or else we must assume that adjunct-specific properties can vary between languages and are acquired by the speaker.

3.2.2 Empirical patterns

The classic syntactic accounts of adjuncts make very clear predictions for the extractability of adjunct clauses – as long as the clause is not selected by the verb it will be an island, either because it is merged late, or not properly governed, etc. These types of accounts will henceforth be referred to as *categorical*. While there is a clear majority of examples of illicit extraction from adjunct clauses in the literature, several counterexamples that require explanation have been adduced. The following sections will focus on such counterexamples to the expected pattern that adjuncts are categorical islands. First, I will present some of the relevant empirical patterns, and then review some non-categorical approaches that have been proposed as a way of handling these patterns, i.e., approaches that do not predict that all adjunct clauses will be islands and see whether they can explain the counterexamples.

Non-finite adjunct clauses

It has been observed that there is a difference between finite and non-finite adjunct clauses with regards to extraction, such that non-finite adjunct clauses might allow extraction (see e.g., Truswell 2007, 2011; Dal Farra 2010). Some examples are presented in (36) from both English and Italian, which are marked by the authors as acceptable:

(36) Acceptable extraction

a. **What** did John drive Mary crazy [whistling _]? (English; Truswell, 2011: 30)

b. ***Quale ragazza*** *Gianni è* *partito [senza salutare _]?*
which girl Gianni is left without greeting
'Which girl did John leave without greeting?' (Italian; Dal Farra 2020: 53)

The categorical approaches do not distinguish between finite and non-finite adjunct clauses regarding how they are integrated with the main clause, and thus, cannot explain that extraction from certain non-finite adjunct clauses is acceptable. Truswell (2007, 2011), presenting a semantic condition on adjunct clauses, assumes that a tense operator blocks movement from all finite adjuncts, but that non-finite adjuncts, which are not stipulated by Truswell (2007, 2011) to have a tense operator, in principle can be extracted from given that they meet certain semantic criteria.

Specifically, Truswell (2011: 157) proposes the semantic criterion *the Single Event Grouping Condition* (SEGC) for non-finite adjunct clauses, which states that:

(37) *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.

To be construed as an event grouping is dependent on two criteria – (i) there is spatiotemporal overlap between every two events; (ii) a maximum of one (maximal) event is agentive. In (36a) there is only one agentive event – *whistling* – and there is spatiotemporal overlap between the *whistling* and the event of *driving Mary crazy*. In (38a) both the event in the adjunct clause (*whistling*) and the event in the main clause (*work*) are agentive, which explains why extraction is ruled out. In (38b) there is not spatiotemporal overlap between *grinning manically* and *meeting* – the natural interpretation of *since* is that the meeting precedes the manic grinning.

(38) Unacceptable extraction

a. ***What** does John work [whistling _]?
b. ***Who** has John been grinning manically [since meeting _]?

(Truswell, 2011: 38, 142)

Interestingly, Truswell (2011: 43) predicts that there will be variation between speakers in the acceptability of forming *wh*-dependencies into non-finite adjuncts, as the construal of two events depends on both context and the individual's life experiences. The acceptability will reflect the *perceived* relations between the specific events and will depend on "world knowledge and the interlocutor's creative ability to perceive links between subevents" (2011: 124). As such, acceptability will naturally be subject to inter- and intra-speaker variation (2011: 123).

Truswell (2007, 2011) assumes a categorical distinction between finite and non-finite adjunct clauses, but that there are fluid boundaries between acceptable and unacceptable extraction within the group of non-finite adjunct clauses. Dal Farra (2020: 108) finds a similar pattern. She finds a statistically significant effect of tense for adjunct clauses in Italian, but that the effect of tense disappears for certain adjunct clause types (as extraction is bad regardless of tense). Similarly, Tanaka (2020) finds that both quantifier raising and *wh*-dependency formation are sensitive to non-finite adjunct island constraints in English. However, in support of Truswell's SEGC, she also finds differences between the adjunct clause types tested: bare gerunds and prepositional phrases with "during" (i.e., activating a spatio-temporal overlap reading) receives higher scores in a *wh*-dependency compared to *after*-clauses (2020: 167-168).

For Mainland Scandinavian languages, however, it seems that tense is not a crucial criterion for extraction. Müller (2019) finds that finiteness does not matter for Swedish adjunct clauses in the same way as Truswell (2011) assumes that it does for English.[21] Kush et al.'s (2019), results strongly suggest that finiteness should not matter for topicalization or relativization in Norwegian either. There is no mechanism within Truswell's SEGC to allow for variation in tense between languages.[22]

Finite adjunct clauses

Truswell (2007, 2011) presents a new way to account for the AC and shows that different explanations for non-finite and finite adjunct clauses is necessary. However, he maintains the old assumption that extraction from finite adjunct clauses is categorically unacceptable, regardless of the semantic relationship between events in the adjunct and matrix clause. As such, for finite adjunct clauses, the SEGC is as categorical as the traditional, syntactic explanations. That there are examples that extraction from finite adjunct clauses is acceptable would therefore be quite surprising.

However, we *do* find such counterexamples across many languages:

(40) Acceptable extraction: finite adjunct clauses

a. %This is **the watch** that I got upset [when I lost _] (English; Truswell, 2011: 175, fn.1)

b. **The person** who I would kill myself [if I couldn't marry _] is Jane.
(English; Deane, 1991: 29)

c. ***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: 115)

[21] Kohrt et al. (2020) find evidence that speakers do not posit gaps inside non-finite adjunct clauses in English. Using electroencephalography (EEG), the authors investigate the processing profiles of filler-gap dependencies and find that adjuncts deviate from the profile associated with licit filler-gap dependencies. The authors interpret this to mean that non-finite adjunct clauses are islands in English such that speakers do not *actively* anticipate gaps inside the adjunct clause. Instead, Kohrt et al. (2020) argue that they find evidence of bottom-up "inactive" dependency formation in adjunct clauses.

[22] Ernst (2022) argues that given the empirical evidence that not all languages are sensitive to the distinction in tense, a parametrization of the SEGC is required. He suggests, here explained using the terminology provided by Truswell, that only some languages require a single-event reading for extraction to be possible, while others allow extraction with a multiple-event reading. Crucially, Ernst argues that a tense operator does not itself block extraction.

d. ***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)

e. ***Na*** ***tian*** *de* *xi* *wo* *[yinwei* *probu* *xihuan e]*
that day MM play I because no like
mu *jian* *xiuxi* *de* *shihou zou* *le*
curtain between rest MM time leave ASP
'The play of that day, I left during the time of break because I didn't like (it)'
(Chinese; Liejiong 1990:461)

The examples in (40) provide evidence that there are cases where extraction out of a finite adjunct clause is accepted. Interestingly, it seems that this is not a special characteristic of one language, but that many different languages potentially can allow such extraction. Evidence such as this is highly problematic for categorical approaches to adjunct islands, and for Truswell's (2007, 2011) similarly categorical prediction for finite adjunct clauses.

The data points in (40) are supported by emerging experimental evidence showing that extraction from finite adjunct clauses is not always unacceptable. In general, the experimental evidence for English supports the prediction made by categorical approaches that finite adjuncts do not allow extraction (Sprouse et al. 2011; Sprouse et al. 2013; see also Villata & Tabor 2022). However, there is evidence that there are differences between *dependency types*. Sprouse et al. (2016) find a distinction between *rc*-dependencies and *wh*-dependencies for English finite *if*-adjunct clauses. They find a classic *island effect* for *wh*-dependencies but fail to find an island effect for *rc*-dependencies. By island effect I mean the effect islands have on acceptability which (i) arises when the tail of a long-distance dependency is inside an island structure, (ii) the effect cannot be explained by any other property of the construction (Sprouse & Villata 2021: 229).[23] Similarly, Kush et al. (2018) find a classic island effect of forming *wh*-dependencies into finite *om* 'if'-adjuncts in Norwegian and only a very small island effect of forming a topicalization dependency into finite *om* 'if'-adjuncts. This indicates that differences between dependency types might be relevant for adjunct extraction. On a constraint-based approach, this is highly surprising as all A'-dependencies are typically assumed to involve the same operations and should not yield distinct results.

[23] This is a simplified definition of *island effect.* For a detailed explanation, please see Section 4.2.

There is also initial evidence that there are differences between *adjunct clause types*. Bondevik (2018), which provides the starting point for the current thesis, provides evidence that there are differences in the size of the island effect between conditional *om* 'if', causal *fordi* 'because' and habitual *når* 'when' in Norwegian topicalization dependencies. Paper 1 elaborates on and extends this work. Müller (2019) informally investigates adjunct clause types in Swedish and finds that there are differences in the average judgments such that some adjunct clause types yield high average acceptability judgments, while others yield low judgments. She concludes that there are differences between adjunct clause types in Swedish and that we need a theory that can explain that certain adjunct clause types might allow extraction.

That finite adjunct clauses which have long maintained a position in the literature as a classic island can be extracted from is highly problematic for the CED, and for the related categorical approaches. As these approaches are based on a basic syntactic distinction between adjuncts and complements, there is no obvious way to maintain the category "adjunct", while also allowing extraction. As such, evidence that finite adjuncts are not always islands strongly suggests that categorical approaches must be altered. In addition, finding that there are differences between dependency types indicates that the distinction between "adjunct" and "complement" is too coarse-grained to properly explain the pattern. Such differences imply that there must be some property of the embedded clauses that interact in some way that is relevant for islandhood with the properties of each of the dependency types. Consequently, we need islandhood to be derived from properties that are sensitive to such interactions. This is also the case if we find differences between adjunct clause types – unless we find that certain adjunct clauses are not adjuncts after all, the theory must necessarily be sensitive enough to distinguish between adjunct clauses that have different properties, be that syntactic, semantic or pragmatic.

3.2.3 Non-categorical explanations of extractions from finite adjunct clauses

The above sections discuss and describe constraint-based approaches to islands. More generally, these can be referred to as derivational locality constraints on A'-movement (Boeckx 2012). A derivational locality condition constrains the structure-building component such that it cannot derive a structure, i.e., the Subjacency Condition. The inability to derive the structure is what causes an island violation. The recurring problem for the current derivational approaches to islands is that they over-generalize patterns of extraction, either by over-generalizing the extent to which extraction from finite adjunct clauses is licit (the Subjacency Condition, the base version of the PIC) or illicit (Late-Merge, Pair-Merge etc.). Subsequently, we should explore less rigid (non-categorical) approaches to islands.

Syntax-based: representational approach to islands

One such approach is Relativized Minimality (Rizzi 1990), which is a constraint-based representational locality condition, as opposed to a derivational locality condition. A representational locality condition sees island constraints as occurring from conditions on the output, i.e., the derivation is complete, but constraints apply to the final syntactic representation.[24] Relativized Minimality sprung from the Government and Binding theory which explains locality constraints as conditions on antecedent government of traces (e.g., Huang's 1982 CED). Specifically, Chomsky (1986a) assumes that a head will block traces from being governed in both a head- and an antecedent-chain (asymmetric intervention). Thus, in (41) below, so long as Z is a head, it will block government between X and Y.

(41) X … Z … Y …

Rizzi (1990: 1-2) refers to this view as "Rigid Minimality" and proposes a relativized version of minimality ("Relativized Minimality") in which an intervenor can only intervene if it is of *the same type* as the chain, in other words, the blocking of an intervening governor will be relative to the nature of the government relation involved. Thus, in (41), only if Z is a potential governor of some kind for Y will Z block government of the same kind from X (Rizzi 1990: 2). Thus, a head will only block government of a head, and only an antecedent will block government of an antecedent (symmetric intervention).

Relativized Minimality (RM) in terms of Rizzi (1990) constrains government of traces. As traces are introduced in the structure after movement occurs, it is clear that RM constrains the *output* of cyclic operations. This is also true within the Minimalist Program. In MP movement is considered to be a type of merge, and as opposed to a moving element leaving a trace, movement is re-conceptualized as there being two copies of the same element. A chain is formed by a word having two separate functions. When a word is semantically selected (θ-role) (external merge) and provided scope-discourse semantic functions (internal merge), the two positions form a chain (Rizzi 2006: 101). In a post-government framework, Rizzi (2001) updates the terms used in the 1990-definition of RM to the following:

[24] For a brief discussion of *where* and *when* RM is evaluated, see Section 5.3.

(42) Definition of Relativized Minimality

> "Y cannot be related to X if Z intervenes and Z has certain characteristics in common with X. So, in order to be related to X, Y must be in a minimal configuration with X, where Minimality is relativized to the nature of the structural relation to be established" (Rizzi, 2001: 89).

I interpret "be related" as chain-formation. Dispensing with government and types of government for determining minimality, the MP determines minimality based on featural overlap between elements in the chain. Thus, the MP version of RM is a featural approach to RM, henceforth referred to as featural RM (fRM, Villata et al. 2016), which states that no dependency can be formed between the filler and the gap if there is an element that matches the features of the filler and the gap and that is c-commanded by the filler and itself c-commands the gap (Rizzi 2013). Starke (2001) further develops this framework and postulates that, given that there are several features that are relevant for fRM, extraction is acceptable if there is an additional feature on the mover that the intervener does not share.

To illustrate how fRM works, *wh*-islands constitute an good example case. *Wh*-islands have been proposed to allow certain elements to move out. Under derivational approaches, any *wh*-element should occupy Spec-CP, which blocks the escape hatch and makes the *wh*-clause an island. However, this constraint is too strong. It has been shown that certain elements can move out from a *wh*-island as in (43a).

(43)

a. ??[$_{CP}$ **Which problem**$_i$ [$_{C'}$ do [$_{TP}$ you wonder [$_{CP}$ how$_j$ [$_{C'}$ [$_{TP}$ to solve _ $_i$ _$_j$]]]]]]?
b. *[$_{CP}$ **How**$_j$ [$_{C}$ do [$_{TP}$ you wonder [$_{CP}$ which problem$_i$ [$_{C'}$ [$_{TP}$ to solve _ $_i$ _$_j$]]]]]]?

(Rizzi 1990: 4)

As is evident from (43), all embedded Spec-CPs are filled, which means that the derivational approaches will rule out any extraction from the embedded *wh*-clauses. fRM can provide an explanation for the extraction pattern in (43). In (43a) a complex *wh*-element *which problem* is moved across an intervener *how*, which is a bare *wh*-element. While in (43b), the bare *wh*-element is moved across a complex *wh*-intervener. Assuming that *wh*-words have criterial features [+Q], i.e., features which trigger movement to a position "dedicated to some scope-discourse interpretive property" (Rizzi & Shlonsky 2007: 116), and that nouns have a [+N]

feature, then a bare *wh*-element will have the featural make-up [+Q] and a complex *wh*-element [+Q, +N]. Thus, (43) can be analyzed as in (44).

(44) Examples in (43) illustrated with feature calculus

a. ??<**Which problem**> do you wonder how to solve <which problem>?
[+Q, +N] [+Q] [+Q, +N]

b. *<**How**> do you wonder which problem to solve <how>?
[+Q] [+Q, +N] [+Q] (Rizzi 1990: 4)

In (43a) and (44a) there is partial overlap between the intervener and moving element such that the moving element is featurally richer than the intervener. Such a featural calculus will according to fRM yield acceptable extraction (Starke 2001). In (43b) and (44b), however, the intervener have all the same features that the moving element has ([+Q]) which leads to ungrammaticality.[25]

fRM is also flexible enough to allow for gradience in acceptability. Rizzi (2018) sketches a system in which the degree of similarity between the intervener and the mover determines the degree of acceptability. Similarity is defined as featural overlap – the less features in common, the less similar two constituents are. Thus, when features do not match, dependencies are fully acceptable; when they partially match, dependencies are deemed partially acceptable; but when features fully match, dependencies are fully unacceptable. Rizzi (2018: 347) outlines a distinctness hierarchy (the arrow illustrates the degree of acceptability) which is illustrated in (45):

(45) Distinctness hierarchy

- X and Z have nothing in common: disjunction
- X has some features distinguishing it from Z, but also some features in common with Z
- X and Z have all relevant features in common

This provides additional explanation of the extraction patterns in (43) and (44). The distinctness hierarchy in (45) predicts that since the intervener in (44a) shares one feature with the moving element there is partial featural overlap which leads to a slight decrease in acceptability. As

[25] See Paper 3 for a more detailed overview of fRM.

(43a) is marked as '??', meaning that it is not a fully acceptable sentence, this is exactly what is predicted by (45).

Villata et al. (2016) test the predictions of fRM in a formal judgment study. In compliance with the predictions of fRM, Villata et al. (2016) find that patterns that yield full overlap of features yield the lowest acceptability judgments, patterns that yield partial overlap yield mid-range acceptability scores and finally patterns where there is no overlap yield the higest acceptability scores.

Other non-categorical approaches

Other syntactic, non-categorical approaches have also been suggested to specifically explain instances of licit extraction from adjuncts. An approach that does not consider the timing of Merge or the type of Merge, is the height-based account of adjunct islands (Brown 2015a,b; Müller 2019 building on Heageman 2012). Within the height-based approaches, the assumption that closely integrated clauses allow extraction, while more peripheral ones do not is maintained from the CED (see e.g., Müller 2019, Dal Farra 2020). Degree of integration is here determined by the structural closeness of the embedding verb and the adjunction site. Thus, it is not *how* adjuncts are introduced in the sentence, but *where* that determines how closely integrated they are with the main clause. Some adjuncts are assumed to be attached high in the clause, modifying CP, while other are attached lower in the clause, modifying TP or VP (Haegeman 2012: 165). The high adverbial clauses have a peripheral status and are less integrated in the clause, i.e., they are attached at the periphery of the clause. The low adverbial clauses are attached low in the clause and, thus, obtain a central status and are more integrated in the clause. Specifically, the assumption is that if the adjunct is adjoined in the VP-domain (low) it is assumed to be semantically integrated with the event/state expressed by the verb (see Haegeman 2012: 7), i.e., semantically close to the verb. Here semantic closeness overlaps with structural closeness. Müller (2019) shows that in cases where the adjunct clause can be classified as a central adverbial clause (i.e., merged low in the structure) *and* there is a coherence relation between the adjunct and the matrix clause (i.e., a semantic condition) extraction might be allowed in Swedish.

Brown (2015a,b) proposes that extraction will be impossible if the adjunct merges with a phase head (*v*P), but that it will be possible if it merges with VP. Brown (2015b) argues that the semantic integration of verb and adjunct correlates with possible extraction – semantic integration of the verb and the adjunct is only possible when the adjunct is merged low in the structure, i.e., below the *v*P phase edge, and this is also the only place of Merge from which extraction is possible without violating minimalist assumptions about locality and movement.

There are also approaches that attempt to capture the AC in terms of general constraints on discourse: Erteschik-Shir's (1973) *Dominance Hypothesis,* Goldberg's (2006) *Backgrounded Constructions are Islands* (BCI) and Abeillé et al.'s (2020) *Focus Background Constraint* (FBC). What these approaches have in common is that they do not emphasize the structural difference between adjuncts and complements, thus, they are not forced to assume a categorical distinction between adjuncts and complements. Instead, discourse-based approaches to islands see island effects as originating from infelicities in the discourse. More specifically, it is proposed that forming an A'-dependency into a domain that does not have a prominent discourse function in the sentence will be infelicitous and that this will yield a decrease in acceptability akin to an island effect.

For instance, under the BCI (Goldberg 2006, Ambridge & Goldberg 2008) all backgrounded constructions are predicted to be islands. Backgrounded elements are constituents that are neither the primary topic nor part of the potential focus domain (Goldberg 2006: 130). Goldberg (2006) shows that certain adjuncts that are not backgrounded allow extraction.

(46) **Who** did she travel to Memphis [ADJUNCT in order to see _]? → EXTRACTION OK

Goldberg (2006: 145) argues that the adjunct clause in (46) is not backgrounded on the basis of the following negation test:

(47) Negation test

She didn't travel to Memphis in order to see Elvis

→ the travel was not done to see Elvis, i.e., seeing Elvis was not the purpose of the trip

= *in order to see Elvis* is not backgrounded

(Goldberg 2006: 145)

Goldberg (2006) writes that the sentence in (47) can be used to imply that the travelling was done, but not in order to meet Elvis. If negation can scope into the adjunct clause, the adjunct clause is *not* backgrounded. As such, Goldberg (2006) shows that the BCI can provide an explanation for why some adjunct clauses, but not all, allow extraction.

A very recent discourse-based approach is *the focus-background conflict* (FBC) constraint presented in Abeillé et al. (2020). This builds on the previous discourse-based approaches, but also considers the discourse function of the dependency that is formed. Abeillé

et al.'s (2020) proposal is that the island effect that is observed with certain domains is a result of a *discourse clash*, such that the discourse function of the domain where the gap belongs does not match the function of the filler.

(48) *Focus-background conflict constraint*:

A focused element should not be part of a backgrounded constituent.

(Abeillé et al. 2020: 3)

Their specific case study is on subject islands in *rc*-dependencies. Abeillé et al. (2020: 19) assume subjects to be backgrounded, unless they specifically receive prosodic stress. They therefore predict subjects to be islands for all dependencies in which the filler is focalized. It is assumed that *wh*-question formation and topicalization are such dependencies, whereas relativization is assumed to be an operation that does not focalize the filler. Accordingly, the FBC constraint predicts *rc*-dependencies into backgrounded subjects to be acceptable.

Overall, for English, Abeillé et al. (2020) find relativization from subjects to be more acceptable than from objects, once preposition stranding is controlled for.[26] This pattern is also found for French. In addition, strengthening the evidence in favor of the FBC constraint, *wh*-dependencies yield quite opposite results: a clear preference for *wh*-dependencies into objects over subjects is found in both languages. Thus, Abeillé et al. (2020) find cross-construction variation *and* cross-linguistic generalizations.

The FBC is defined in quite general terms, which allows the validity of this constraint to be tested in a variety of languages, island domains and dependencies. Specifically, adjuncts that are presupposed are typically assumed to be backgrounded (Goldberg 2006: 134). Under the BCI, such adjuncts are thus islands. Under the FBC-constraint, however, backgrounded adjuncts will be islands only in *wh*- and top-dependencies, but not in *rc*-dependencies. Thus, backgrounded adjuncts should not be islands for relativization. Such generalizations are supposed to be cross-linguistically valid, given that the FBC constraint is proposed to be a universal constraint on discourse felicity (Abeillé et al. 2020: 3).

[26] Abeillé et al. (2020) argue that the effect of preposition stranding in English is the driving force behind the large subject island effects found for English *rc*-dependencies in Sprouse et al. (2016). Abeillé et al. (2020) find that an *rc*-dependency into a subject where the preposition is stranded is much less acceptable than for objects, where this type of extraction is accepted. Unfortunately, this difference between subjects and objects is not explored in more detail in Abeillé et al. (2020).

Kobzeva et al. (2022) test the predictions made by the FBC. In a large-scale study testing the difference between *wh*- and *demonstrative rc*-dependencies for four different island types; subjects, embedded questions, adjunct *if*-clauses, and relative clauses, they find that their results cannot be explained by the FBC. First of all, they find that extraction from subject clauses is categorically illicit in Norwegian, regardless of the type of dependency. Next, they find that, while adjunct clauses display the predicted difference between *wh*- and *rc*-dependencies, the distribution of scores is not compatible with the FBCs predictions. They argue that their data suggests that other semantic/pragmatic factors that go beyond a distinction between backgrounded and focused constituents are needed to explain cross-dependency differences.

3.3 Not all empirical patterns are predicted

In summary, there is a discrepancy between the theoretical predictions and the empirical data – the categorical approaches predict that adjunct clauses always will be islands, but the empirical data suggests that some languages do allow extraction from certain adjuncts (both finite and non-finite). Such a discrepancy is a major issue for the categorical accounts of the AC. As I have shown in this chapter, the categorical accounts of the AC predict a ban on extraction from finite adjunct clauses based on the assumption that adjuncts are less integrated, more peripheral, on a separate plane – spatially and temporally – compared to complements. In these theories, it is not possible for an element to be an adjunct (i.e., not selected by the verb) *and* open for extraction. That we do, however, see such patterns, both for finite and non-finite adjunct clauses, provides powerful counterexamples to established theories.

Specifically, there is initial evidence that there are differences between dependency types and adjunct clause types. To account for such patterns a more fine-grained theory than what has been previously suggested is required. There are some non-categorical approaches that potentially allow for more fine-grained explanation. However, one of the major obstacles for theory development in this area is that we do not yet have a clear grasp of the empirical landscape, e.g., are all adjunct clauses the same for long-distance A'-movement?; are all A'-dependency types the same for adjunct clauses? Accordingly, we first need to investigate the finer details of the empirical landscape rigorously and, in more detail, before potentially abandoning or amending theoretical approaches.

4 Methodological issues

The current chapter will provide a discussion of the issue of collecting linguistic data and describe in detail one approach to collecting linguistic data (the "Sprouse-design"). The chapter is organized as follows. In Section 4.1, I discuss linguistic data in general, which type of linguistic data are useful for investigating island phenomena and the methodology for collecting linguistic judgments. In the next section, the specific experimental design which has been used to gather data for the project will be described in detail, before limitations and other considerations are tackled in Section 4.3.

4.1 Linguistic data

The big-picture aim of the project is to determine whether the grammar allows Norwegian speakers to form certain filler-gap dependencies into certain adjunct clauses or not. This is however a difficult task as we cannot observe the grammar directly. "The grammar" is the mental state of the speaker, which constitutes stable knowledge about language, as opposed to linguistic behavior. This distinction between linguistic knowledge and linguistic behavior goes back to Chomsky (1965). Chomsky (1965: 4) makes a distinction between linguistic *competence* and *performance*, where linguistic competence refers to the "speaker-hearer's knowledge about his language", and performance "the actual use of language in concrete situations". Linguistic performance will vary as a factor of context, psychological state of discourse participants (e.g., stress, anxiety) etc., while linguistic competence is the stable and integrated knowledge that we have of our languages. The theoretical linguist is interested in the linguistic competence of the language user and must therefore abstract away from the data of performance to the underlying system of rules that make up the grammar (Chomsky 1965: 4). In fact, linguistic theory is interested in mental processes that "are far beyond the level of actual or even potential consciousness" (Chomsky 1965: 8-9). Thus, the theoretical linguist is faced with a "black box" problem. Linguistic competence is a mental state it cannot be directly observed – we can observe the input and the output, but not how this is organized or computed in the language faculty. Chomsky (1959/1967) argues in his famous critique of Skinner and the behaviorist view of language, that as opposed to making linguistic behavior the principal object of study, linguistic behavior simply provides the only data we have available that has the potential to inform us of the "black box". Thus, the main object of study was shifted from input/output to the mental state.[27]

[27] This shift in the object of linguistic study has been referred to as a revolution within the field of linguistics (Åfarli 2000: 194; Joseph 2010; Hårstad, Lohndal & Mæhlum 2017: 84).

Accordingly, we must contend with studying the output of the mental state, namely performance data. And the choice of linguistic performance data depends on the phenomenon under investigation. The two most common sources of linguistic performance data are corpus searches (actual language use) or judgment data (perception of language). A corpus search has the advantage that all cases encountered are cases of actual language use, either written or spoken. A problem is that the absence of any given construction in the corpus does not mean that this construction is not attested in other corpora or produced by speakers (see Henry 2005 and Gross 2021 for a discussion of this). Subsequently, a corpus search is not necessarily informative, unless we have reason to believe that the constructions we are interested in frequently occur in the corpus. Island violating structures are theorized not to occur naturally in language. Thus, a corpus search might not provide us with any new knowledge, unless we do find instances of island violations.[28]

Judgment data have proved to be an important source of linguistic data. Linguistic judgment data constitute data on speakers' *perceptions* of linguistic content. It is often compared to other types of perception data such as perception of light, pain, temperature etc. (see e.g. Schütze & Sprouse 2014; Gross 2021). Speakers are typically asked to what extent they perceive the stimuli to be acceptable or to compare their perception of two sentences. Thus, the linguist will present linguistic content to speakers and ask the speaker to judge the perceived *acceptability* of the linguistic content. The assumption is that measuring perceived acceptability will disclose knowledge about *grammaticality*, i.e., the mental representation of language. Underlying mental representations have since Chomsky (1957) been considered the only stable predictor of linguistic behavior within the generative framework (see e.g., MacCorquodale 1970 for a critical discussion of this assumption).

Providing intuitions about data is a form of linguistic performance[29] that hinges on memory, contextual cues, fatigue, personal preferences etc. Linguists now typically emphasize

[28] See Kush, Sant & Strætkvern (2021) and Müller & Eggers (2022) for two studies investigating island phenomena through corpus searches. Corpus searches together with acceptability judgment studies might also provide interesting findings. For instance, Adger (2013) argues that the absence of parasitic gaps in corpora together with findings that speakers of English agree on the acceptability of parasitic gaps indicates that island repair with parasitic gaps must be a constraint specific to language, as opposed to a general cognitive constraint.

[29] Goodall (2021: 8) discusses what type of behavior is at play when speakers are "reacting" to linguistic data and writes that the term "judgement" may be a good description, except that "it suggests that the process involves protracted and conscious deliberation on whether the sentence is acceptable or not, whereas in practice, the process appears to be virtually instantaneous".

that they are testing "acceptability" as opposed to "grammaticality", as the term "grammaticality" implicates that we can directly access the mental grammar. This is considered to be too strong an assumption about the linguistic judgment data that we have access to. Goodall (2021: 8) emphasizes that on this view of linguistic data "grammaticality" and "acceptability" are not synonyms but that "(un)grammaticality" is simply one of the influences of "(un)acceptability" as a sentence that is well-formed (according to principles of the grammar) may be considered unacceptable (e.g., due to parsing difficulties). Moreover, Goodall (2021: 8-9) assumes that acceptability is a "sensation" that is equally as gradient as other sensations. This point emphasizes the importance of not equating acceptability and grammaticality. The gradience uncovered in acceptability judgments may be caused by competence-independent factors which are not relevant for the topic under the investigation. Thus, gradience in judgments does not equal a gradient mental grammar.

Judgment data can either be collected "informally" by asking friends and colleagues about their perception of certain data, or "formally" in a controlled experimental setting. Informal judgment data collection typically involves that the linguist asks a few speakers of a given variety to provide judgment on data that probes the phenomenon under investigation. In practice, this can involve the linguist consulting their own judgment and other expert subjects or a few "lay" subjects. Informal judgment data collection is much less time- and resource consuming compared to formal judgment collection. Accordingly, if informal judgment collection is sufficient as a method for the phenomenon under investigation, this is a good choice (Gross 2021).

Informal judgment data has up until recently been the typically method for studying island phenomena. As island violations have been theorized to be clearly unacceptable, the effect of an island violation should be easily detectable with informal judgment data, i.e., performance factors should not interfere to a large extent. However, it seems that some island phenomena do not provide straight-forward intuitions, which challenges the assumed validity of such methodological approaches to islands. To take one example from research on Norwegian: Bermingrud (1979) and Faarlund (1992) mark the equivalent sentences in *bokmål* and *nynorsk* (the two co-existing standardized written varieties of Norwegian) as unacceptable and acceptable, respectively.

(49) Ambiguous informal judgments

a. ***Det*** *blir* *han* *sint* *[når* *jeg* *sier* *_].* (Bokmål)
That becomes he angry when I say
'That, he gets mad when I say' (Bermingrud 1979: 89)

b. ***Det*** *blir* *han* *sint* *[når* *eg* *seier _].* (Nynorsk)
That becomes he angry when I say
'That he gets angry when I say' (Faarlund, 1992: 115)

There is no reason that the two standardized varieties of Norwegian should differ in this specific respect. Each of the authors has simply chosen to illustrate examples in different varieties, and the difference in acceptability between Bermingrud (1979) and Faarlund (1992) should be taken as different judgments of an equivalent structure. This example illustrates that informal judgment data are not necessarily sufficient for studying island phenomena.

4.1.2 Formal judgment data for studying islands

One interesting tool that has been further developed in recent years are formal acceptability judgment experiments. In carefully controlled experiments, gathering acceptability judgments is formalized to provide more reliable data. Large participant groups are included and tested on a large set of variations of the same structure to reduce the chance of participant or test sentence confounds. This is particularly informative in cases where variation between participant groups or inter-structure variation has been proposed. In addition, inferences made on the basis of large samples should be more reliable as it allows us to aggregate over individual differences, infelicities, performance-issues or semantic oddness of one or two test sentences etc.

Additionally, formal acceptability judgments are provided by people who have little or no prior formal linguistic training – "lay" subjects. This is a strength because, unlike informal judgments which typically make use of colleagues, participants do not have any preconceptions about the sentences or more seriously, personal research agendas related to the outcome of the judgment test. Thus, lay subjects can in theory provide unbiased judgments. At the same time, these participants might provide their judgments on properties of the test sentence that are unrelated to the phenomenon under investigation – perhaps they find the context to be unusual or dislike the wording etc. In fact, it is highly likely that the participants are not aware of the phenomenon that is being tested as formal acceptability judgment experiments take measures to prevent participants from detecting which phenomenon is being investigated. Subsequently, they will judge each test sentence as a whole and will not provide judgments on the phenomenon that is in question. This means that the acceptability judgments of individual sentences must be

aggregated over and that single scores for test sentences must not be given much theoretical weight as they might be the result of performance-related issues etc.[30]

Thus, strikingly different from typical informal intuition data, no one sentence can be thought of as an ideal representative of that type. Instead, there must be several sentences that act as representatives of any given type (Cowart 1997: 47). In other words, researchers must be cautious about how to interpret acceptability judgments on experimental stimuli gathered from lay subjects. Schütze (2020: 190) argues that "finding that subjects' ratings on a set of experimental stimuli do not align with the published judgments of linguists does not necessarily represent a genuine data discrepancy". As such, he cautions against "falsificationism" based on evidence that participants do not provide the same judgments as expert linguists do on specific sentences (2020: 193). Following the view of linguistic judgment data argued for in Schütze (2020), researchers must maintain a cautious approach to judgment data as there are many sources that can influence judgment that do not rely on the mental grammar. In a study investigating this problem, Schütze (2020) finds that there are a number of cases where participants disagree with linguists' judgments, but where participants disclose in a post-judgment-study-interview that there were reasons beyond "grammaticality" that caused them to rate a given sentence well or badly. The reasons why ungrammatical sentences were rated highly were mis-parsing or leaving out short words (such as "I" or "to"). The reasons grammatical sentences were given low ratings ranged from un-natural word choices ("I would never say this", "it sounds funny" etc.) to mis-parsing and similar issues.[31]

There are measures that researchers can take to make formal acceptability judgments studies as reliable as possible. For instance, Cowart (1997) argues that filler sentences must be included in the experiment to best avoid that participants identify the phenomenon under investigation. Specifically, Cowart (1997) argues that there should be an equal amount of test sentences to filler sentences (= balanced design). The design of the test sentences is also crucial. Schütze (2020) points out that careful control and construction of items is necessary to ensure that participants respond to the intended distinctions, e.g., island vs. no-island. It is therefore

[30] Lay subjects may also be less confident in performing linguistic judgments.

[31] It might be the case that lay subjects (on average) are not as influenced by other factors as we might think. Sprouse & Almeida (2017) find that informally gathered judgments by and large provide the same results as experimentally gathered judgments (though this might also indicate that expert subjects might be equally influenced by such factors).

important to keep the test sentences as similar as possible, only varying on the crucial distinction that is relevant for the study. Particularly, Cowart (1997: 11) writes:

> "[…] the important point about judgments of acceptability is not that all of the variation in judgments is controlled by the syntactic properties of the target sentences but that some of it is. […] Thoughtful research design and modern statistical techniques make it quite practicable to design experiments that, through their structure, isolate those components of variation in judgments that are attributable to syntactic manipulations, regardless of what other factors may be contributing to overall variation in judgments in a particular experiment".

Here Cowart (1997) advocates for an approach to formal acceptability judgments which builds on other scientific frameworks. For instance, in a physiological study where the effect of a given endurance training intervention is investigated, researchers will test participants on the same conditions, but train participants differently (e.g., different intervention groups, control groups etc.). The important point will be to have comparable participant groups (age, endurance level, weight, height) at the beginning of the study (pre-testing) and to *compare* the different groups after the study (post-testing). The comparison between groups on the post-tests makes up the data. Any statistically significant difference between groups will be attributable to the training intervention, as the groups are comparable on all other factors, though there might be a myriad of other factors that also contribute to differences between participants (e.g., form, sleep patterns, nutrition etc.) that are not controlled for. These factors will, however, be distributed more or less equally across the different intervention groups (according to statistical logic). In applying such a methodological approach to syntactic intuition data, Cowart writes that instead of suppressing extra-syntactic influences, the aim is to "control these influences by trying to spread them as uniformly as possible across all the sentence types to be tested" (Cowart 1997: 47). Just like a physiological researcher wants to have as uniform groups as possible before training intervention is applied to control for factors such as age, weight, height, nutritional habits etc., the linguist wants extra-syntactic influences such as length, frequency of words, and context to be uniformly distributed across test sentences. Thus, if the goal is to test the effect of an island violation, sentences that violate an island barrier and sentences that do not will be compared, but the goal in setting up the design must be that it is only the island violation that differs between sentences – "all other lexical, syntactic, semantic, pragmatic and miscellaneous psychological factors being uniformly distributed across the two sentences" (Cowart 1997: 47).

4.2 The Sprouse-design

There are many additional considerations that must be taken to design a reliable acceptability judgment experiment. These will be illustrated through a description of the "Sprouse-design" which is used to collect data for the papers in the current thesis. The Sprouse-design is an established design for studying islands and has been used across many different languages and island types (Sprouse et al. 2011, 2012a, 2016; Kush et al. 2018, 2019; Paper 1; Kobzeva et al. 2022; Paper 2). With this design we can compute a quantitative measure of any potential decrease in acceptability of an island violating sentence compared to non-island violating sentences. This is the so-called *island effect*, i.e., the effect on acceptability that an island violation has. Thus, utilizing this design allows for an easy comparison with previous research as it provides a measure of the effect of an island that can easily be compared across experiments using the same design.

4.2.1 The philosophy of the design

Throughout all experiments in Papers 1 and 2, the experimental design popularized by Sprouse (2007) was used. The quantitative measure of an island effect that this design provides makes it easy to compare island effects across clause types, dependency types and languages. In addition, the design has the advantage that for each island type it can (in theory) separate between two hypotheses regarding the source of island effects. There are two main competing hypotheses for why island domains do not allow extraction (see Sprouse et al. 2012a; Sprouse et al. 2013 and Hofmeister, Casasanto & Sag 2013 for a debate, and see also Phillips 2013a,b for further discussion) – the constraint-based hypothesis (Ross 1967; Chomsky 1973, 1977, 1986a, 1995, 2000, 2001; Huang 1982, see e.g. Boeckx 2008a or den Dikken & Lahne 2013 for a reivew), which I discussed at length in Section 3, and the processing hypothesis (see e.g., Deane 1991; Kluender & Kutas 1993; Hofmeister & Sag 2010; Hofmeister et al. 2013). As discussed above, within the constraint-based hypothesis, islands violations arise because there are certain general (and universal) constraints on language that are being violated. The processing-based hypothesis, also referred to as the reductionist view, on the other hand, seeks to eliminate as many innate grammatical constraints on language as possible and sees island violations as arising from an overload on the processing device. Hofmeister & Sag (2010: 379) assume that for filler-gap dependencies this means that processing filler-gap dependencies will be less costly if there are fewer processing demands "along the filler-gap path". If reductionist approaches are correct, the constraint-based approaches' endeavor to find underlying grammatical constraints is futile. In addition, the reductionist view yields the most minimal

grammar, and therefore we should only propose underlying grammatical constraints in cases where it is clear that a reductionist view cannot explain the extraction patterns.

Finding that a given island domain yields a low acceptability score is not informative in isolation as this could indicate that the processing cost of this sentence is too great (i.e., no grammatical constraint is activated), or that there are constraints on forming A'-dependencies into this domain (i.e., a grammatical constraint must be activated). In other words, judgments on a single sentence type does not help us to distinguish between the two competing hypotheses. The Sprouse-design, however, can potentially help us to separate between these hypotheses. An island violation consists of two main factors: (i) there is a domain that is different from a regular declarative clause; (ii) *and* there is long-distance movement. Crucially, these two factors are only problematic in combination, as is illustrated by the fact that both sentences in (50) are acceptable.

(50)

a. Specific type of domain
   I believe [NP the claim [CP that Otto was wearing the hat]] => Complex NP
b. Long-distance movement
   I like **the hat** that I believe that Otto was wearing __ . => Relativization

On the reductionist view, however, it is assumed that sentences containing a complex structure like (50a) or a long-distance dependency like (50b) are costly to process and that acceptability of these sentences will be slightly lower than for a sentence like (51):

(51) I believe that Otto was wearing a hat.

In (51) there is a matrix clause and an embedded declarative clause introduced by "that". There is no domain which has been postulated to constitute an island, and second, there is no filler-gap dependency. (51) is subsequently assumed to be less costly to process and accordingly more acceptable than both sentences in (50). Finally, acceptability breaks down when there are two costly structures in the very same sentence, like in an island violating sentence as in (52):

(52) ***The hat** that I believe the claim that Otto was wearing _.

Thus, Sprouse (2007) postulates that comparing four sentences like (51), (50a), (50b), and (52) will yield a step-wise decline in acceptability under the simplest implementation of the reductionist approach. The sentence with an embedded declarative (51), being the cheapest structure to process, will yield the highest acceptability scores, the sentences in (50a-b) are slightly more costly and consequently slightly less acceptable than (51), while (52) the costliest and less acceptable than (50a and b). This hypothesized relationship between sentences (50, 51 and 52) is illustrated in Figure 1.

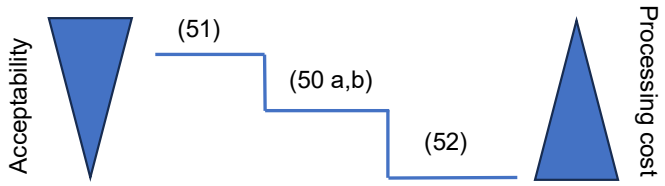


Figure 1: Step-wise decline in acceptability and a step-wise increase in processing cost illustrated with the examples provided above.

The reductionist view takes island violations to be caused by the added sum of processing cost that each of the factors constitutes. Again under the simplest implementation of the reductionist view, adding the decrease in acceptability of a long-distance filler-gap dependency ((51) vs. (50a)) to the decrease in acceptability of processing a complex domain ((51) – (50b) will provide an accurate prediction of the acceptability of an island violating sentence (52).

Imagining that the sentences in (51) and (50a,b) above were rated as follows, we predict that if the reduced acceptability is caused by a processing effect, the acceptability of the island violating sentence should be a factor of the difference between (51) and each of the two sentences in (50), as illustrated in Table 3 and (53) below:

Table 3: Overview of imagined acceptability rating for each test sentence, including an overview of sentence structures that are hypothesized to be costly to process.

Rating	*Test sentences*	*Sentence structures*
5	(a) I believe that Otto was wearing a hat	
4	(b) I like **the hat** that I believe that Otto was wearing _	*Relativization*
4	(c) I believe [NP the claim [CP that Otto was wearing the hat]]	*Complex NP*
3	(d) **The hat** that I believe [NP the claim [CP that Otto was wearing _]]	*Relativization* + *Complex NP*

(53) Calculations Table 3

Step 1: Effect of relativization (a – b): 5 – 4 = 1

Step 2: Effect of Complex domain (a – c): 5 – 4 = 1

Step 3: Effect of both (a – (Step 1 + Step 2)): 5 – (1 + 1) = 3

All steps together: 5 – ((5 – 4) + (5 – 4)) = 3

The judgment pattern in Table 3 together with the calculations in (53) provide an example where the acceptability of each condition in isolation directly predicts the acceptability of conditions combined. Thus, in Table 3, the acceptability of test sentence (d) (conditions combined) is predicated on the difference in acceptability between (a) and (b) and (a) and (c). Thus, if there linear addivity, the judgment of test sentence (d) is the same as the judgment of test sentence (a) minus the effects of the (b) and (c) test sentences, as in (54).

(54) Judgment pattern linear addivity: a – (a – b + a – c) = d

The judgment pattern can be visualized in an interaction plot as in Figure 2, and if the judgments are as provided in Table 3 above, there is a *linear additivity effect* as illustrated under "No island effect".

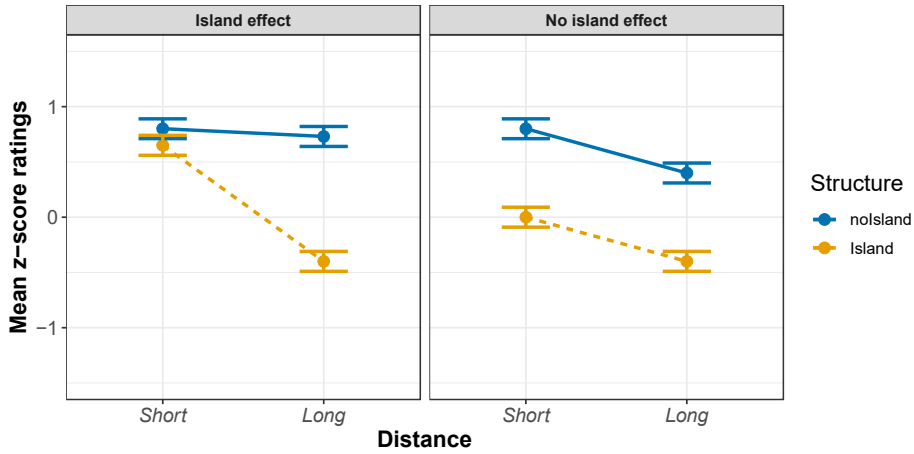


Figure 2: Visualization of predicted acceptability judgment patterns – interaction effect ("island effect") vs. no interaction effect ("no-island effect")

Sprouse (2007) terms such a pattern a "no-island effect", meaning that if there is linear additivity, there is no independent island constraint at work, but "simply" a processing effect.[32]

If we see a decrease in acceptability on the island violating sentence that is bigger than the added sum of each of the factors, this is termed an "island effect". With such judgment patterns, we can assume that there is some additional element that causes an added decrease, which is typically explained in terms of abstract grammatical constraints on filler-gap dependency formation (Sprouse & Villata 2021: 234), i.e., a syntactic (the CED – Huang 1982) or semantic constraint (the SEGC – Truswell 2007; 2011). Such a pattern provides a *superadditive effect,* which visually will look different compared to the linear additivity effect and where the acceptability judgment of the island violating sentence cannot be predicted based on the other conditions as in (55).

(55) Judgment pattern superadditivity: $a - (a - b + a - c) \neq d$

[32] Importantly, this assumption about what a processing effect will look like has been questioned in recent years. See e.g., Sprouse & Villata (2021) and Kim (2021) for a recent discussion of this issue. See also Hofmeister, et al. (2012) for general criticism of the assumptions about processing effects in the Sprouse-design. Several superadditive processing effects have been proposed, which, if correct, would mean that the two hypotheses cannot be distinguished. This shows that we should not take the effect at face-value, but that the source of the interaction effect must be determined based on the whole experiment.

Imagining that the sentences in Table 3 are instead rated as in Table 4, we get a different judgment pattern:

Table 4: Different acceptability rating for each test sentence compared to Table 3

Rating	*Test sentences*	*Sentence structures*
5	(a) I believe that Otto was wearing a hat	
4	(b) I like **the hat** that I believe that Otto was wearing _	*Relativization*
4	(c) I believe [$_{NP}$ the claim [$_{CP}$ that Otto was wearing the hat]]	*Complex NP*
1	(d) **The hat** that I believe [$_{NP}$ the claim [$_{CP}$ that Otto was wearing _]]	*Relativization* + *Complex NP*

(56) Calculations Table 4

$5 - ((5 - 4) + (5 - 4)) = 3$

$d \neq 3$

If ratings are as illustrated in Table 4, the acceptability of (d) cannot be predicated on the difference between (a) and (b) and (a) and (c), as shown in (56). This judgment pattern yields a super-additive interaction effect, visualized in Figure 2 under "Island effect".

As such, the Sprouse-design is more informative than single acceptability judgments of island violating structures as it provides a diagnostic for distinguishing between two hypotheses. This is a major advantage of formal acceptability judgment studies.

4.2.2 Details of the design

The design is typically set up as a 2x2 ('two-by-two') factorial design. A 2x2 factorial design is a common design for studying phenomena with different components that potentially interact. Each of the two factors can be carefully controlled, and the effect contributed by each effect measured. In this particular design, the effect of crossing the two factors STRUCTURE and DISTANCE is of interest. The effect of each of the factors must be measured separately. Each of the factors have two levels, or more directly, two different realizations. For *structure*, the design compares sentences where there is what has been claimed to be an island structure [+ISLAND] or a declarative *that*-clause [-ISLAND]. For *distance*, sentences where there is either a long-distance filler-gap dependency [LONG] or a short-distance filler-gap dependency [SHORT]. This yields four test conditions that are different realizations of the two factors illustrated in Table 5.

Table 5: Overview of factors in the test sentences in Table 3

Factors	*Test sentences*
[-ISLAND][SHORT]	(a) I believe that Otto was wearing a hat
[-ISLAND][LONG]	(b) I like **the hat** that I believe that Otto was wearing _
[+ISLAND][SHORT]	(c) I believe the claim that Otto was wearing the hat
[+ISLAND][LONG]	(d) **The hat** that I believe the claim that Otto was wearing _

As discussed above, an island violation requires there to be both an island domain present, and a filler-gap dependency that is formed into it. Thus, Table 5 (d) should be the only ungrammatical condition under the constraint-based approach as this is the only realization of the factors *structure* and *distance* that exactly fits this pattern. The design allows controlling for and measuring the effect of each of the variables involved in causing an island violation separately: the effect of STRUCTURE is obtained by comparing the *island* vs. *no-island* conditions and the effect of DISTANCE by comparing the *long-distance* vs. *short-distance* conditions. Accordingly, the design makes it possible to measure the effect that the factor DISTANCE has on acceptability separately for each of the structures – subtracting Table 5 (a) – (b) isolates the effect of DISTANCE in a non-island, while subtracting Table 5 (c) – (d) isolates the effect of DISTANCE in an island. Similarly, the design makes it possible to isolate the effect of STRUCTURE for each level of DISTANCE.

Following Sprouse et al. (2016) there are three ways to identify an island effect in this design (which is the main goal of an experiment using this design): (i) visually inspecting the interaction patterns (linear additivity vs. super-additivity); (ii) calculating *differences-in-differences* scores (DD-scores) which provide a numerical identification procedure for the size of an island effect and finally; (iii) fitting linear mixed effects models in a statistical procedure. Before I go into the details of these procedures, I will first describe each of the experiments shortly and how the data was collected.

The current project comprises five separate experiments – three of which are reported in Paper 1 and two in Paper 2. The experiments are very similar in that they are set up with the Sprouse-design and test extraction from at most three different finite adjunct clause types, in addition to the same two control clause types. I will distinguish between the experiments by naming them according to the distinguishing features. The experiments in Paper 1 test extraction from finite *fordi* 'because', *når* 'when' and *om* 'if' by way of topicalization, while the experiments in Paper 2 test extraction from the same adjunct clause types by relativization. I will describe the topicalization and the relativization experiments separately.

For paper 1 investigating topicalization dependencies into three different adjunct clause types – *fordi* 'because', *når* 'when' and *om* 'if', I ran three experiments using the Sprouse-design. Experiment Top1 tested 8 items for each adjunct clause type and 8 items for the control island types – subject islands and complement *om* 'whether'-clauses. Following-up on surprising findings in Experiment Top1, two experiments were conducted in which participants were exposed to twice as many stimuli to test the robustness of the surprising findings in Experiment Top1. In Experiment Top2a, *fordi* 'because' and *om* 'if' adjunct clauses were tested with 16 different items, in addition to 8 items for each control island type. In Experiment Top2b, *når* 'when' and *om* 'if' were tested with 16 different items (the same 16 for *om* 'if' as tested in Experiment Top2a), in addition to the same 8 items for each control island type. For all three experiments, items were distributed across 4 lists in a Latin Square fashion and pseudo-randomized within each list.

Paper 2 investigates relativization dependencies into the same three adjunct clause types, and here two experiments were run. Experiment Rel1 was run in the exact same way as the Top2ab experiments where three adjunct clause types were tested together with two control conditions. Items were distributed across 4 lists in a Latin Square fashion. Experiment Rel2, however, also used the Sprouse-design, but with a slight alteration. Here only *om* 'if' was tested together with the same fillers used in Experiment Rel1. Each participant encountered each test sentence (i.e., no Latin Square design) and saw all test sentences in the exact same order, i.e., no counterbalancing of material within the experiment. This makes it possible to also control for potential variation between participants, items and order of exposure. This means that participants were exposed to many test sentences with the exact same structure and to all variations of the same item, which raises saturation concerns. Cowart (1997: 50) warns that participants' judgment of the second encounter of a token-set (i.e., item) will be influenced by the first encounter. This is problematic for analyses as it adds a second confound in the interpretation of differences between sentences, namely *order*. However, importantly, we do not see any effects of order in Experiment Rel2. It might be that the number of fillers (balanced design) or the obscure nature of island phenomena protected against order effects.

The data for all experiments were collected using online acceptability judgment tests with native Norwegian speakers. Participants were recruited through online platforms, e.g., social media or through posts on university-internal platforms (Paper 1: Experiment Top1 and Top2a, Top2b, Paper 2: Experiment Rel2). Participants were recruited through Prolific for Paper 2: Experiment Rel1.

Experiments in Paper 1 were hosted on IbexFarm (Drummond 2012), and experiments in Paper 2 were run on JATOS with JsPsych (de Leeuw 2015). For all experiments, participants accessed the experiment via a link, which led them to a consent form, followed by instructions[33] and an unmarked practice phase, before accessing the main experiment. Figures 3 and 4 illustrate what the experiment phase looked like in each case:

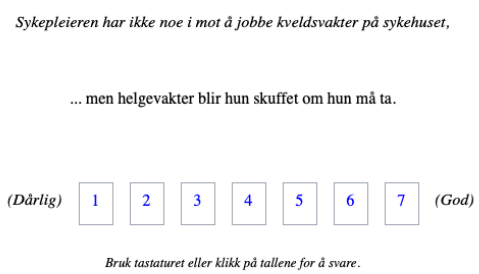


Figure 3: Sentence presentation IbexFarm – preamble in italics, test sentence provided below. A full 1-7 Likert Scale is presented below the test sentence.

The instructions *bruk tastaturet eller klikk på tallene for å svare* 'use the keyboard or click on the numbers to provide your response' was included on every page of the experiment. The translation of the test sentence in the example in Figure 3 is as follows:

(57) Translation of test sentence shown in Figure 3

Preamble:

Sykepleieren	*har*	*ikke*	*noe*	*i*	*mot*	*å*	*jobbe*
nurse.DEF	has	not	something	in	against	to	work
kveldsvakter	*på*	*sykehuset.*					
nightshifts	at	hostpital.DEF					

'The nurse does not mind working night shifts at the hospital'

[33] Paper 2, Section 3.4, details the instructions provided both in Paper 1 and Paper 2.

Test sentence:

... men	***helgevakter***	*blir*	*hun*	*skuffet*	*om*	*hun*
... but	weekend.shifts	becomes	she	disappointed	if	she
må	*ta.*					
must	take					

'but… weekend shifts she is disappointed if she has to take'.

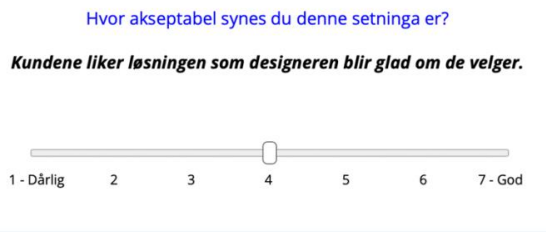


Figure 4: Sentence presentation Jatos – test sentence provided in bold, and the task-instruction in blue. A full 1-7 Likert Scale is presented below as a gliding scale, but the arrow will only stop at each number.

The question *hvor akseptabel synes du denne setninga er?* 'how acceptable do you think this sentence is' appears on each page of the experiment. The translation of the test sentence in the example in Figure 4 is as follows:

(58) Translation of test sentence shown in Figure 4

Kundene	*liker*	***løsningen***	*som*	*designeren*	*blir*	
customers.DEF	like	solution.DEF	that	designer.DEF	becomes	
glad	*om*	*de*	*velger.*			
happy	if	they	choose			

'The clients like the floor plan that the designer will be happy if they choose'

In addition, as the figures illustrate, for each experiment judgments were provided on a 1-7 Likert Scale – as is typical with this design (Sprouse 2007; Sprouse et al. 2012a; Sprouse et al 2016; Kush et al. 2018, 2019). 1 was given as *dårlig* 'bad' and 7 as *god* 'good', which Marty et al. (2020) define as a full Likert Scale. Sentences were presented each at a time (single presentation). Marty et al. (2020: 3) find that a full Likert Scale provides higher effect detection rates with single presentation than a non-labelled Likert Scale, and that this is especially true

for medium sized effects. Sprouse & Almeida (2017) argue that for small effect sizes, the sample size of participants must be above 100 if each participant only sees 1 item per experimental condition for the statistical power to be strong enough to detect an effect with a Likert Scale test. If, however, the number of items per experimental condition is increased, there is a much larger chance of detecting small effect sizes with a Likert Scale test. For these reasons, we provided the participants with full Likert Scales and several items for each experimental condition.

Raw Likert Scale scores are typically linearly transformed to increase comparability between trails and to avoid scale biases. Z-scoring is a type of linear transformation that 'standardizes' the score. First, the raw scores are centered such that each data point is expressed as a measure of how much it is above or below the mean. Next, the centered variable is divided by the standard deviation of the sample (Winter 2020: 86-87). Thus, a z-score expresses "how many standard deviations the score is away from the mean" (Winter 2020: 88). Linear transformations do not alter the shape of the data but change the units of the x-axis (Winter 2020: 88). However, importantly, z-scoring within participants, which is how z-scoring is used within the Sprouse-design, does affect the relationship between data points as it is not a linear transformation across the entire experiment. Thus, a z-score of 0.1 for one participant does not correspond to the same raw score as z=0.1 does for another. Instead, a very positive effect of z-score transforming the data, as pointed out in Featherston (2005: 681) is that "the different scales that the individual subject adopted for themselves" are unified.

4.2.3 Data analysis

The first step in the data analysis, after visualizing the data, is to look for interaction effects (which here correspond to island effects). This is done by fitting a linear mixed effects model on the data set. A linear model is a null hypothesis statistical test that provides an estimation of a value for a given data point given a value of another data point – i.e., "modelling y as a function of x" (Winter 2020: 69). In the current context, we are interested in what happens with the judgment value (y) when the syntactic value is "island" vs. "declarative sentence". Mixed models can handle datasets where there is a dependency between data points (Winter 2020: 233) such as in the current experiments where participants provide several data points, the same sentence is rated many times (by different participants) etc. In effect, a linear mixed effects model fits varying slopes and intercepts (= random effects) for the dependencies that it has been informed about (e.g., varying slopes by participant). Thus, the linear regression is able to take into account varying slopes and intercepts for the data points that are not independent (Winter

2020: 234). What the model *does* is not to estimate parameters per participant/sentence, but instead to provide an estimation of the variation around the specified random effect (2020: 238).

The linear mixed effects models fitted on the data specified DISTANCE and STRUCTURE and their interaction as fixed effects, and by-participant and by-item varying slopes and intercepts. Thus, for each level of both DISTANCE and STRUCTURE the model estimates an average z-score while taking into account variation between participants and items. In addition, the model returns the probability that the null hypothesis is true. The null hypothesis for a linear mixed effects model with an interaction term is that there is no effect of either factor and no interaction effect of the factors. If the model returns a p-value below 0.05, the null hypothesis is falsified. Thus, the model both provides an estimation of values, while also providing a falsification test for whether or not there is an island effect (=interaction effect).

The *size* of any potential island (i.e., interaction) effect can be calculated as a DD-score. The DD-score gives us a measure of how "much greater the effect of an island structure is in a long-distance dependency sentence than in a sentence with a local dependency" (Sprouse et al., 2012a: 92). Thus, we are interested in comparing the difference between the condition where there is a local dependency and no island domain ("short, no-island") to the condition where there is an island violation ("long, island"). The DD-score can be stated mathematically as ([+ISLAND, LONG] – [–ISLAND, SHORT]) = ([–ISLAND, SHORT] – [–ISLAND, LONG]) + ([–ISLAND, SHORT] – [+ISLAND, SHORT] + X, where X is the island effect (Sprouse & Villata 2021: 230). This can be algebraically rearranged as the following equation in order to calculate X:

(59) Formula for calculating a DD-score

X =

([+ISLAND, LONG] – [–ISLAND, LONG]) – ([–ISLAND, SHORT] – [+ISLAND, SHORT])

(Sprouse & Villata 2021: 230)

The way the DD-score is calculated is: (i) calculate difference scores between the levels of STRUCTURE for each level of DISTANCE [–ISLAND, LONG] – [+ISLAND, LONG] and [–ISLAND, SHORT] – [+ISLAND, SHORT]; (ii) calculate the difference between the difference scores for each level of STRUCTURE – which essentially gives us the difference between the long and short conditions (DISTANCE), once the difference between [±ISLAND] has been calculated.

The DD-score provides a measure that can easily be compared between languages (Sprouse et al. 2016), types of islands (Sprouse et al. 2016; Kush et al. 2018, 2019; Paper 1; Paper 2), types of dependencies (Sprouse et al. 2016; Kobzeva et al. 2022) and types of

participant groups (Kush & Dahl 2020) as the 2x2 factorial subtraction logic ensures that an island effect is an isolated measure of the differences between the main effects (STRUCTURE and DISTANCE).[34]

4.2.4 Other considerations

The types of statistical analyses that can be performed on the data set depend on the type of data. Z-scoring the Likert Scale ratings is a standard procedure with the Sprouse-design (see e.g., Sprouse et al. 2012a; Sprouse et al. 2016; Kush et al. 2018, 2019). However, there might be a caveat which concerns the data type. Z-scoring the Likert Scale ratings assumes that the scale is interval. This means that we assume the distance between 1-2 to be the same as the distance between 4-5. We do not have a clear understanding of how participants use the scale in acceptability ratings; thus, it might not be true of a 1-7 Likert Scale that participants treat the scale as truly interval, though finding evidence for or against this assumption is difficult. There are, however, alternative ways of analyzing the Likert Scale ratings which do not assume interval data. Instead, taking the data as ordinal allows us to run very similar tests as with z-scored ratings without assuming the Likert Scale to be a true interval.

Ratings are considered ordinal if there is a natural ordering between them such that one naturally comes before the others (for instance, a Likert Scale with "strongly disagree" vs. "disagree") (Levshina 2015: 17). The responses differ in their order, but the size of the difference between the points on the scale is unknown. As such, an advantage of treating the responses as ordinal is that we do not need to make assumptions about how participants view the scale. Taking the data to be ordinal is the conservative choice. Interval data constitute a higher-level data type and allow more statistical tests than ordinal data.

To investigate this more thoroughly, I conducted a similar analysis for Experiment Rel2*om* on the raw responses, assuming the data to be ordinal, as I did for the z-scored ratings. The results of the analyses on the z-scored ratings are detailed in Paper 2, while the results of the ordinal analyses are provided here. In order to fit a regression model on ordinal data, the *probability* for a test sentence of receiving a certain response level on the Likert scale was used. Specifically, we look at the *log odds ratio*. The odds ratio provides a measure of probability – an odds ratio of 1 is the same as saying "the chance is 1 to 1" and equals 50% chance of something occurring. The log odds ratio is a logarithmically transformed odds ratio transforming the data to a continuous measure of probability (Winter 2020: 203). A (natural) log odds ratio of 0 equals an odds ratio of 1.

[34] If we obtain a negative DD-score, we can interpret this as a *reverse island effect* (Sprouse et al. 2011).

To conduct the data analysis, I followed the procedure detailed by Barlaz (2022) for ordinal logistic regression in R (R Core Team 2021): data visualizations followed by fitting an ordinal logistic regression proportional odds model on the data. Data visualizations were created with ggplot2 (Wickham 2016). Broadly speaking, the results obtained from the z-scored raw ratings were replicated with the raw scores, which shows that the analyses treating the data as interval might be as reliable as the analyses treating the data as ordinal. Below a detailed overview of the ordinal analysis for Experiment Rel2*om* follows.

I ran an ordinal logistic regression in R using the clmm()-function from the ordinal package (Christensen 2019). The clmm()-function allows fitting main effects and their interactions in addition to random effects in the model. I fitted a mixed ordinal logistic regression model with DISTANCE × STRUCTURE as the main effects, and by-item and by-subject varying intercepts and slopes.[35] The model returned a significant interaction effect ($p = 0.0027$). It also showed a significant main effect of DISTANCE ("short" vs. "long") ($p > 0.001$), and a main effect of STRUCTURE ("no-island" vs. "island") ($p > 0.001$). Estimating the probability (0-100%) for each response level by condition using the ggpredict()-function from the ggeffects package (Lüdecke 2018) we see that the probability of receiving a certain score increases by the increase in response level for the three conditions "short, no-island", "short, island", "long, no-island", i.e., higher responses are estimated as more probable. For the "long, island"-condition we also see an increase in probability by response level, but here there is a decrease in probability at "6" and even more at "7". The estimated probabilities are plotted in Figure 5.

[35] The model's syntax: clmm(response~structure * distance + (1 + structure*distance |item) + (1 + structure*distance|subject),data = df, link = "logit").

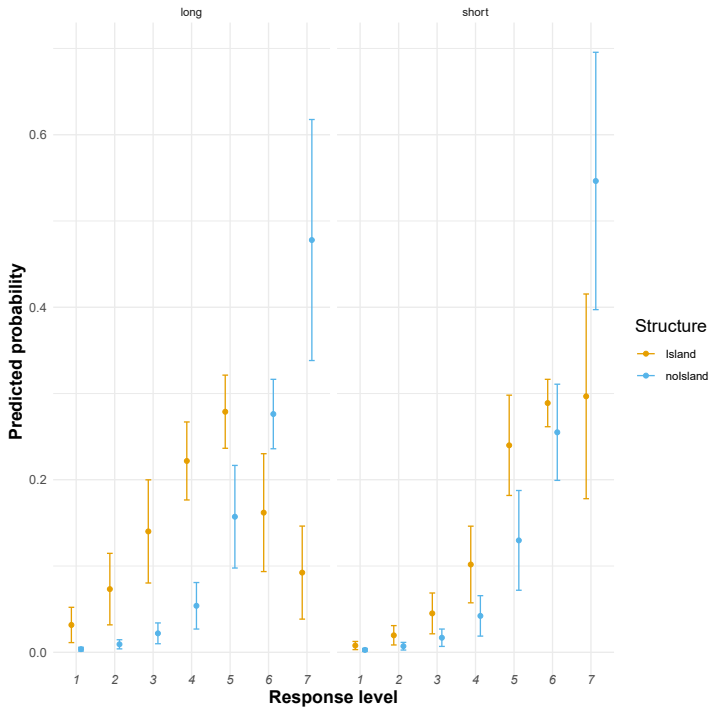


Figure 5: The dot represents the predicted probability and the error bars represent the range of the confidence level. We see a wider range of confidence levels, which we can interpret as more variation, for the high end of the Likert Scale.

There is an interaction effect in the predicted probabilities: the decrease in probability of receiving a "7" or "6" is lowest for the "long, island"-condition. It is interesting to see that it does not decrease until after response level "5". This indicates that although there is an interaction effect, the probability of a relatively high judgment is high, which is also visible in the z-scored interaction plots in Paper 2.

Going beyond the procedure detailed by Barlaz (2022), the response level on the "long, island"-condition was fitted by participant in a simple logistic regression with ordinal proportional odds using the clm()-function from the ordinal package (Christensen 2019) in R (R Core Team 2021). This returns an estimation of the log odds ratio of each response level per participant. Next, I calculated the predicted probability of each response level by participant based on the log odds ratio. We see that there is variation between participants. This is plotted in Figure 6.

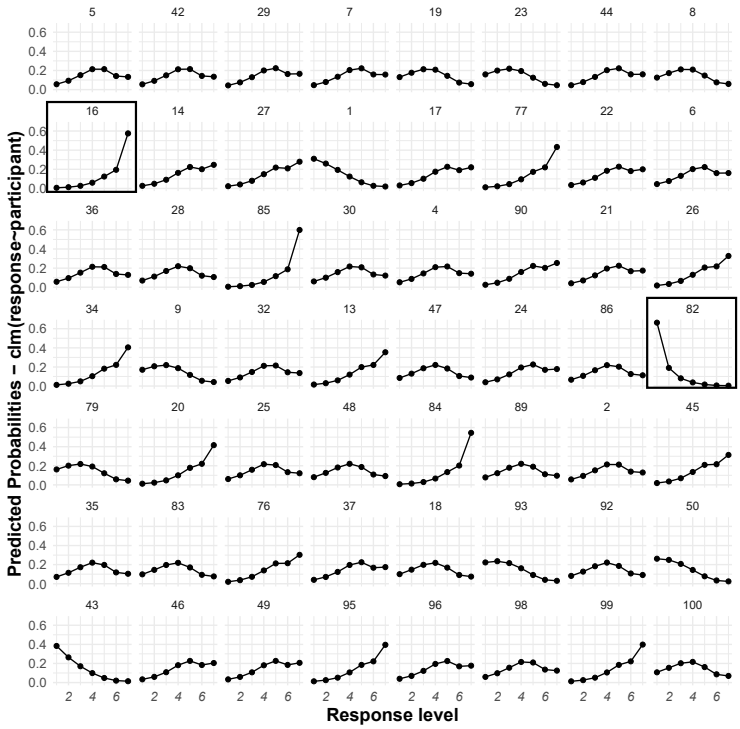


Figure 6: Predicted probability of response levels by participant are plotted for the "long, island"-condition. Participants "16" and "82" are highlighted for easy comparison.

The plot reveals that participants have used the scale differently – whether this reflects that speakers have different judgments or if this is a task effect is difficult to ascertain from the data. To illustrate, for participants "82" and "16" a raw response of "3" probably means very different things. Z-scoring the raw responses allows us to capture this difference. For participant "82" a raw response of 3 will yield a very high z-score, whereas a very low one for "16". Thus, it seems that z-score transforming the data prior to data analysis is indeed a well-founded procedure.

Seeing similar results for the raw responses and for the z-scored responses indicates that the main conclusions do not change depending on the response type used for analysis. The raw data can provide a better understanding of what participants are really doing when they judge the sentences. Investigating the raw responses is a good "reality check" when used as a supplement to the z-scored ratings. First of all, it provides more insight into the absolute values such that an island-violating condition that mostly receives "7" cannot be maintained to be an island-violation. However, the raw responses introduce a second confound of insecurity in the

data. When the predicted probability of receiving a "5" on the "long, island"-condition is high, we do not know whether this is (1) because all participants collectively find this intermediate rating most representative of the acceptability of these types of sentences, or (2) because participants use the scale differently such that a rating of "5" is considered high for some participants, but low for others. Accordingly, we could see a pattern where generally rejected items are rated "5" by participants who tend to use only the high-end of the scale, and generally accepted items are rated "5" by participants who tend to only use the lower end of the scale. Naturally, following such a pattern, the most likely response level overall will be "5". Z-scoring the data controls for these scale biases across participants. The two scenarios sketched would yield different z-scores for "5"; (1) would yield a z-score of 0 if "5" is the mean judgment for that participant; (2) would yield a positive z-score for participants who tend to use only the low-end of the scale and a negative z-score for participants who tend to use only the high-end of the scale. Accordingly, the z-scored ratings provide a measure of the acceptability of a given condition relative to the acceptability of other conditions. As the z-scores are calculated by participant, we only see the relative ordering of sentences per participant. Thus, if a test sentence receives mainly positive z-scores, we know that this test sentence is rated as more acceptable than the mean of the other sentences included in the experiment across participants, regardless of what the raw scores are. We can say that relative to other sentences in the experiment this sentence is accepted.

Still, this is not without problems. The z-score ratings are dependent on the ratings of every score included in the experiment. Thus, the z-score is experiment-dependent and cannot be compared across experiments. Raw scores can more easily be compared across experiments and languages. In addition, the interpretation of a given z-score rating might sometimes be quite far removed from the interpretation of the raw response. An example from Experiment Rel2*om* shows that for one participant a raw rating of "5" corresponds to a *z*-score of -0.321. This tells us that a certain test sentence that is encountered is perceived as poorer than another, but it does not tell us much about how this participant perceived the sentence in real terms. A negative z-score is typically taken as a rejection in a balanced design, but it is not obvious that a response of "5" really is a rejection. Accordingly, the z-score is quite far removed from what the participant is really doing when judging the sentences. Then again, we do not know whether the participant is a participant who reserves "5" for unacceptable sentences.

This is exactly why the Sprouse-design is useful – it allows us to diagnose islandhood beyond the mere judgment of a given sentence. As this discussion has shown, the ratings, whether raw responses or z-score transformed responses, are difficult to interpret relative to the

bigger question of acceptance or rejection. The Sprouse-design depends on the judgment of all four conditions within an item to diagnose islandhood.

Nonetheless, as with any design, there are certain limitations that should be discussed. First, the design does not provide an interpretation of differences in effect sizes. The design allows us to interpret a non-island effect as originating from a processing effect and an island effect as an effect of a grammatical constraint. However, several studies, including the papers in the current thesis, report that there are differences in the size of the effect (see e.g., Kush et. 2019). It is a strength of the design that it allows us to uncover such fine-grained differences. Additionally, Papers 1 and 2, Kush et al. (2019), and Kobzeva et al. (2022) report small island effects, and ratings on the "long, island" condition in the acceptable range. Such results are not predicted on any of the approaches, and the design does not provide a possible answer as to the source of such results.

Second, the main basis of the design is that the measure of an island effect is obtained from a comparison with minimally different sentences, as opposed to relying only on judgments of the island violating sentence. In other words, the acceptability judgment of the island violating sentence is not considered informative in isolation. Though this is a well-founded practice as single judgments do not provide information about the source of potential decrease in acceptability, this characteristic of the design also raises potential confounds as to the results. Specifically, it means that lower ratings on any of the other conditions can impact the results to a large extent as the judgments on any of the baseline conditions is considered to be equally informative for the island effect as the judgments of the "long, island"-condition. This can be illustrated with two different scenarios where the ratings of the "long, no-island"-condition is the only difference. In each scenario, described in (60) and illustrated in Figure 7, the judgment pattern under "Island effect" in Figure 2 is taken as the starting point.

(60) Description of two scenarios

a. Scenario A:
   The "long, no-island"-condition is rated as low as the "long, island"-condition. Here we will not see an island effect, but linear additivity.
b. Scenario B:
   The "long, no-island"-condition is rated as high as the "short, no-island"-condition. Here we will see an island effect.

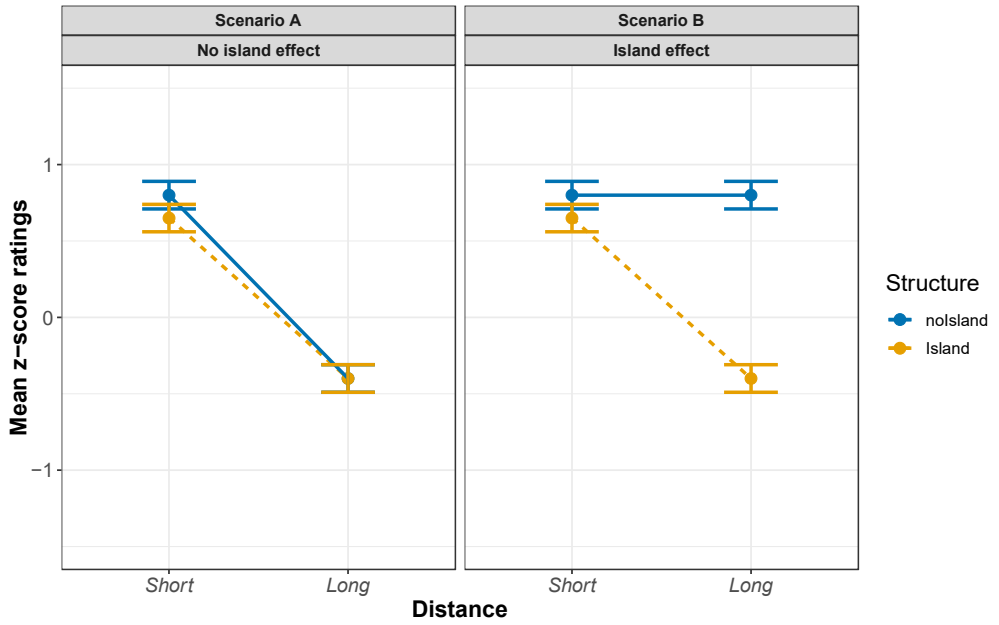


Figure 7. The starting point for each of the described scenarios is the "Island effect" in Figure 2. Each of the scenarios are illustrated on their own under "Scenario A – No island effect" and "Scenario B – Island effect".

The only difference between scenario A and B is the rating of the "long, no-island"-condition. Thus, reporting an island effect vs. linear additivity does not tell us anything about the "long, island" condition in isolation, but tells us something about this condition *in comparison* with the other conditions. A potential confound with this is that there might be island-independent reasons why the "long, no-island"-condition is rated low, which are not controlled for, and consequently, a non-island effect is reported on grounds that are not related to the island violating condition.[36] Additionally, the measure of the island effect is used to theorize about the

[36] As pointed out by a reviewer for Paper 2, the discrepancy in results between Experiment Rel2*om* and Kobzeva et al.'s (2022) results for demonstrative *rc*-dependencies into finite *om* 'if'-clauses in Norwegian could potentially be explained by this aspect of the design. Kobzeva et al. (2022) report linear additivity for *dem rcs* into *om* 'if' in Norwegian, while Paper 2 find a significant, but small island effect for *rc*-dependencies into *om* 'if'. However, an important difference in the results is the difference between the ratings of the "long, no-island"-condition in the two experiments. In Paper 2, the "long, no-island" condition is rated on a par with the "short, no-island" condition, while in Kobzeva et al. (2022) it is rated much lower. The difference in results between the two studies might reflect that there is a general effect of long-distance *dem rc*-dependency formation (i.e., both "long"-conditions are rated low) that is not seen for long-distance *rc*-dependency formation (i.e., only the "long, island"-condition is

"long, island"-condition, which is the condition that we are really investigating, and the basis for this should not be results confounded by island-independent effects at play in any of the other conditions. Doing this will make the design less informative than looking at acceptability scores on the island violating condition alone. For this reason, it is important that studies using this design pay equal attention to the well-formedness of each condition in the test items.

Third, studies using this design have also revealed patterns of variation. Kush et al. (2019) find evidence of between- and within-speaker variation. This is something that the design in its original format does not control for. For this reason, Experiment Rel2*om* in Paper 2 is set up as an alternative to the base version of the Sprouse-design. Experiment Rel2*om* employs the basic properties of the Sprouse-design by testing *om* 'if'-clauses in a 2x2 factorial design, crossing the factors DISTANCE × STRUCTURE, but presents test sentences in a way that controls for participant, item, and order variation. Instead of using a Latin Square Design, all participants see every test sentence, and see them in the exact same order. This allows us to calculate DD-scores for each participant (aggregating over items) and for each test item (aggregating over participants). In case there is variation in the DD-scores by participant, we can assume that this is truly caused by participant variation, and not by the fact that some participants saw some sentences and other participants other sentences etc. Such design alterations have potential confounds and limitations of its own. First, participants are presented with four versions of the same test sentence, which might allow them to identify what is being tested (see Cowart 1997 for a discussion of this). Second, as participants were presented with all test sentences of the same items, they necessarily encountered many test sentences, which meant that, in order to not exhaust participants, we had to limit the experiment to only test *om* 'if', as opposed to testing five different island types in the base version of the Sprouse-design. This means that the experiment is less effective for testing large sets of island types, as is done in Sprouse et al. (2016), Kush et al. (2018, 2019) and Kobzeva et al. (2022). More importantly, Chaves & Putnam (2020) find that participants' judgment of finite adjunct *if*-clauses satiated after repeated exposure to similar test sentences, and this is a potential confound for Experiment Rel2*om*. We do not find, however, any evidence of satiation in Experiment *Rel2om*. This indicates that this procedure can be used, at least when participants are exposed to no more than 16 of the same conditions. Chaves & Putnam (2020) find satiation effects for *if*-clauses with 24 exposures to the same condition.

rated low). It is, however, also possible that there is some additional island-independent confound in one of the experiments for the "long, no-island"-condition that affects ratings.

Finally, Kim (2021) provides an argument of caution against only using the Sprouse-design to investigate island effects. Although finding an interaction effect does not necessarily imply that a syntactic condition is at play, the design in its base format controls for *syntactic* differences between conditions, as opposed to discourse functional differences or semantic differences. An experiment with a design that also captures an island effect in comparative terms, but which tests differences between discourse functional properties or semantic properties should also be possible to conduct, e.g., BACKGROUNDED × MOVEMENT.

4.2.5 Summing up the design

In essence, this design takes advantage of the dual character of island violations: we need to have a specific type of domain and a specific type of filler-gap dependency. An "island" is a type of domain that has been shown not to allow extraction out of it, but an island domain alone is not an ungrammatical structure – we find island domains in many unequivocally grammatical sentences, and similarly, a long-distance dependency is not ungrammatical in and of itself. The design allows us to measure effects on acceptability of each of these factors *alone* and in an interaction.

There are many strengths to this design, most particularly the possibility it provides to compare island effects between dependency types, island types and languages. In addition, though it does not offer a direct interpretation of such effects, it has been shown to be able to uncover fine-grained differences between domains that were previously considered to be the same for extraction. Thus, the design is a very useful tool for investigating islands, but should, as with all research designs, be used with care and the data it provides should only be interpreted on the basis of what the design actually controls for.

5 Main findings

The dissertation addresses the following research questions, as formulated in Section 1.2, repeated here in (61).

(61) Research questions

Empirical research questions

a. Are different adjunct clause types the same for long-distance A'-dependencies in Norwegian?
b. Are different long-distance A'-dependencies the same for different adjunct clause types in Norwegian?

Theoretical research question

c. How can the observed extraction patterns be analyzed formally?

Following the research questions, the aim of the thesis is two-fold: (i) map out the empirical realities of extraction from adjunct islands; (ii) investigate how the observed empirical realities can be analyzed formally. Papers 1 and 2 address the first aim by investigating three different adjunct clause types in two different dependency types. These studies reveal that "adjunct" is not a uniform class with regard to islandhood. Together the studies also show that top- and *rc*-dependencies are the same for the different adjunct clause types. Paper 2 also determines that there is gradience in the acceptability of extracting from conditional *om* 'if'-adjuncts. The empirical reality uncovered by Papers 1 and 2 is surprising under the traditional approaches to adjunct islands and forces a new explanation of adjunct island effects. Such a new explanation is explored in Paper 3, which addresses the second aim of the dissertation. This study argues that adjunct clause types that yield different acceptability scores are qualitatively different island types that are constrained by different locality conditions. Thus, the main finding of the thesis is that adjunct islands behave like a heterogenous group, and that the different adjunct clauses constitute different island types.

This section will first detail the main findings of the three studies, the two empirically directed papers in Section 5.1.1 and the theoretically directed one in 5.1.2. Section 5.2 presents and discusses the new approach that the present dissertation argues for. Theoretical implications of the findings and the proposed account are discussed in Section 5.3. Finally, a comparative view of the main findings is presented in Section 5.4.

5.1 Main findings of the three studies

5.1.1 Empirical findings: Evidence of systematic variation

First of all, the empirical studies show that the three adjunct clauses tested – conditional *om* 'if', causal *fordi* 'because' and habitual *når* 'when' – yield statistically significant island effects. This means that the original generalization made by Huang's (1982) CED is supported by the findings; all adjuncts are islands. In other words, across adjunct clause type, extracting by way of relativization and topicalization yields acceptability scores that are lower than for relativization and topicalization out of declarative complement *at* 'that'-clauses, which are not islands. The current thesis illustrates, however, that not all adjunct clauses are islands *in the same way.* It is not useful to describe conditional *om* 'if' in the same way as causal *fordi* 'because' and habitual *når* 'when' because their extraction patterns are so different that they cannot belong to the same class of "islands". Thus, overall extraction from the three adjunct clauses tested in the current thesis yields lower acceptability judgments than extraction from declarative *at* 'that'-clauses, but the degree to which acceptability is lowered differs considerably between adjunct clause type. Accordingly, the three adjunct clauses should not be treated as a uniform group.

Another interesting generalization is that top- and *rc*-dependencies can be unified. Both dependencies yield the same results for the three adjunct clause types across experiments. A unification of A'-dependencies goes back to Chomsky (1973). However, Sprouse et al. (2016), Abeillé et al. (2020) and Kobzeva et al. (2022) show that there are differences between *wh*-dependencies and *rc*-dependencies, finding extraction by relativization to yield smaller island effects than extraction by *wh*-dependency formation. Neither Sprouse et al. (2016) nor Kobzeva et al. (2022) find island effects for *rc*-dependency formation into conditional *if/om* 'if' in English and Norwegian, respectively. Yet, both studies report island effects of forming *wh*-dependencies into conditional *if/om* 'if'. In a similar vein, Kush et al. (2018) find large island effects of forming *wh*-dependencies, both bare (*what, who*) and complex ones (*which N*), into *om* 'if' in Norwegian. Kush et al. (2019), however, find, in a very similar study, no island effect of forming a top-dependency into conditional *om* 'if' in Norwegian. These findings together with the two current studies reported in Papers 1 and 2, respectively, suggest that top- and *rc*-dependencies form a group to the exclusion of *wh*-dependencies for extraction from conditional *if/om* 'if'. Abeillé et al. (2020) predict that top- and *wh*-dependencies will form a group to the exclusion of *rc*-dependencies, but the current studies show that this is not true for Norwegian. As Abeillé et al.'s (2020) proposal relies heavily on the prediction that *rc*-dependencies will be

different from top- and *wh*-dependencies for all backgrounded clauses, the studies presented in Papers 1 and 2 provide strong counterexamples to Abeillé et al.'s (2020) proposal.

As one of the main findings of the current dissertation is that we must distinguish between adjunct clause type, I will summarize and discuss empirical findings for each adjunct clause type separately below.

Fordi 'because'

Causal *fordi* 'because' clauses were tested in three experiments across the two different dependency types (top- and *rc*-dependencies): ExperimentTop1, ExperimentTop2a and ExperimentRel1. Across experiments, *fordi* 'because' yields large island effects. The effect of extraction out of *fordi* 'because' in both an *rc*- and a top-dependency most closely resembles the effect of extracting out of a complex subject in Norwegian. The complex subject island was included across experiments as a control for classic island effects. Extracting out of a complex subject clause yields large island effects, i.e., strong unacceptability of the island violation condition alone (see Kush et al. 2018, 2019). Accordingly, *fordi* 'because' can best be classified as a classic island.

Following Kush et al. (2019), for each experiment the distribution of the scores was investigated. For *fordi* 'because', I see a narrow distribution around the average judgment across experiments. This indicates that there is little variation between trials for this adjunct clause type. In other words, across items, participants agree that extraction out of *fordi* 'because' is unacceptable in Norwegian.

Fordi 'because' fits the classic pattern captured by the CED (Huang 1982). A DP in object position cannot extract out of the causal adjunct clause. It could suffice to assume that *fordi* 'because' is an island due to its function as an adjunct. However, as will be discussed below, this is not an adequate explanation of the effect as other adjunct clause types show different behaviors.

Når 'when'

Habitual *når* 'when'-clauses were tested in three experiments across the two different dependency types (top- and *rc*-dependencies): ExperimentTop1, ExperimentTop2b and ExperimentRel1. While causal *fordi* 'because' yields classic island effects in both a top- and an *rc*-dependency, the effect of extracting from habitual *når* 'when' is much less easily classified. First, in a top-dependency, *når* 'when' yields scores that on average are comparable to *fordi* 'because'. Yet, the distribution of scores for the "long, island"-condition for *når* 'when' is much wider than for the same condition for *fordi* 'because'. In fact, the scores tend toward being bimodally distributed. This means that there are either-or-judgments of extracting from *når*

'when'. As the average judgment patterns closely together with the average judgment for *fordi* 'because', it is clear that extraction by way of topicalization from *når* 'when' is mostly perceived as unacceptable to participants. However, the bimodality seen in the distribution indicates that on trials where the sentences are not rated as unacceptable, they are instead perceived as acceptable, as opposed to being just slightly less unacceptable.

The variation between trials for *når* 'when' in a top-dependency cannot currently be explained. It is possible that there is variation between participants. Potential variation between participants could be caused by differences in the state of their mental grammars, which would translate as *når* 'when' being an island for some participants but not for others (see Kush et al. 2019 for a discussion of this possibility). Another possibility is that there are differences between test sentences. It could be the case that certain test sentences have a different semantic or pragmatic interpretation than other sentences, and that this difference in interpretation influences extractability. Perhaps the difference in semantics or pragmatics leads to a different parse or makes coercion (see Villata & Tabor 2022) more accessible compared to sentences that receive low acceptability scores. Unfortunately, as participants and test sentences were distributed across four lists, it is impossible to fully disentangle the two potential sources of variation with the Sprouse-design.

In an *rc*-dependency, *når* 'when' yields classic island effects. In fact, *når* 'when' yields scores that are indistinguishable from *fordi* 'because' in a linear mixed effects model. In other words, the model does not rebut the null hypothesis that the two clause types are the same for relativization. In addition, though there is evidence of somewhat more variation between trials for *når* 'when' compared to *fordi* 'because' in an *rc*-dependency, the variation is much smaller compared to the variation seen for *når* 'when' in a top-dependency. This does not necessarily mean, however, that top- and *rc*-dependencies are fundamentally different for *når* 'when' in ways that matter for islandhood. This will be discussed in more detail below in Section 5.

Om 'if'

Conditional *om* 'if'-clauses were tested in all experiments. The findings for *om* 'if' are the most surprising. Across experiments, *om* 'if' yields scores that are markedly different from the other two adjunct clause types: *Om* 'if' yields ratings in the intermediate range below declarative *at* 'that'-clauses, but well above *fordi* 'because'-clauses, which indicates that there is gradience in the acceptability of extraction from adjunct clauses. The surprising results yielded by this clause type warrants a slightly longer discussion as well as a broader discussion of gradience in acceptability and what this means for islandhood.

Om 'if' yields significant island effects both in a top- and an *rc*-dependency. However, the results across experiments clearly show that *om* 'if' is not an island in the same way that *fordi* 'because' and *når* 'when' are. As the results for *når* 'when' are less clear-cut and potentially confounded by factors that are not controlled for in the experiments, I will mainly use *fordi* 'because' as a point of comparison for *om* 'if'. Where *fordi* 'because' yields classic island effects and the average judgment on the "long, island"-condition is well below z=0 (specifically at z=-0.7), *om* 'if' yields small island effects. More specifically, the size of the island effect is small, around half of that of *fordi* 'because', *and* the average judgments on the "long, island"-condition fall around z=0. This does not mean, however, that *om* 'if' is not an island. The linear mixed effects models return significant island effects for *om* 'if' across experiments. Also, the average judgment on the "long, island"-condition for *om* 'if' patterns much lower than for declarative *at* 'that'. Accordingly, the judgments for *om* 'if' fall in an intermediate position between classic islands and non-islands like declarative *at* 'that'-clauses.

The central problem here is to how to *interpret* such intermediate effects, and this is a recurring topic in both Papers 1 and 2. As discussed in detail in Section 3, constraint-based approaches to islands assume a binary division between being an island and not. Consequently, the island effect that *om* 'if' displays which is smaller and different from the classic island effects displayed by *fordi* 'because' is problematic to account for. Before any theoretical account is amended to account for this surprising finding, it should be determined whether the intermediate effect represents the true underlying pattern. There is a possibility that intermediate effects are caused by aggregating over variable judgments, for instance masking underlying variation between participants or items. In other words, an intermediate effect could be an indication that there is underlying variation between trials.

Although there is some variation between trials for *om* 'if' uncovered in Papers 1 and 2, the findings in Paper 2 show that the variation is not meaningful and that the intermediate effect size represents the true underlying pattern. Though the majority of scores pattern above z=0 (i.e., positive z-scores), the distribution of scores on the "long, island"-condition is wide for both top- and *rc*-dependencies, indicating that there is some variation between trials. As for *når* 'when', the source of the variation is impossible to determine in the classic Sprouse-design. A follow-up experiment in Paper 2 was run to control for the obvious sources of between-trial variation, namely between-participant, -item, and -order variation. Investigating each of these, no indication of consistent participant variation, nor any indication of meaningful variation between items is found. Finally, investigating the order of effects, no difference between the first half of the experiment and the second half is found. This indicates that participants did not

e.g., rate sentences better at the beginning of the experiment and worse at later stages, or the other way around.

Consequently, it can be concluded that the intermediate ratings for *om* 'if' represent the true underlying acceptability pattern. In other words, participants seem to consistently perceive extraction by relativization and topicalization as "partially acceptable". Accordingly, it is highly likely that it is something in the underlying system which permits participants to have at minimum a ternary distinction in acceptability – unacceptable (*fordi* 'because' and subject-islands); partially acceptable (*om* 'if') and fully acceptable (declarative *at* 'that').

5.1.2 Theoretical findings: Evidence of two different locality conditions

Having mapped-out a new part of the empirical landscape of extracting out of adjuncts in Norwegian, I will now move on to the second aim of the project, investigating how the observed adjunct island effects can be formally analyzed. As detailed in Section 3, there are several existing approaches to adjunct island effects, which I divide into categorical and non-categorical approaches. The categorical approaches rely on the dichotomy between complements and adjuncts to explain adjunct islands, going back to Huang's (1982) CED. These approaches are categorical in that there is no "intermediate phrasal category" between a complement and an adjunct, and consequently, these approaches cannot integrate intermediate island effects. The consistency with which *om* 'if' differs from *fordi* 'because' in the current experiments provides a very strong argument against the existing categorical approaches in which all clauses that are adjuncts are islands. The non-categorical approaches explain adjunct island effects in terms outside the distinction between complements and adjuncts. These approaches will on the face of it have more power to explain intermediate effects. I will first show the problems the empirical findings of the dissertation pose to the categorical approaches, before I move on to discuss some of the problems for the non-categorical approaches.

The assumption underlying the Sprouse-design is that a quantitatively significant island effect is interpreted as a *grammatical* constraint, as opposed to a *processing* constraint (see e.g., Sprouse 2007 and the debate in the series of papers in Sprouse et al. 2012a > Hofmeister et al. 2012 > Sprouse, Wagers & Phillips 2012b). As I find significant island effects for all adjunct clause types for both *rc*- and top-dependencies, I will assume that this indicates that there is a grammatical constraint that conditions extraction from adjunct clauses in line with the assumptions of the Sprouse-design.

Let us consider two categorical and grammar-based approaches in more detail. Under the Late-Merge approach (Stepanov 2001, 2007), *fordi* 'because', being an adjunct clause, is merged post-cyclically and, consequently, nothing can A'-move out of the clause. One solution

for explaining the intermediate pattern for *om* 'if' would be to postulate that the temporal dimension of merge is ternary. *Om* 'if' does not display the behavior of elements that are merged cyclically (extraction is licit) nor post-cyclically (extraction is illicit). In the MP the assumption is that there are two times at which elements can be merged – cyclically at the time of active derivation, at a time where further movement possible, or post-cyclically, e.g., once the phase is complete, at a time when movement is no longer possible. However, there is no obvious way to postulate an intermediate timing for merge which should correspond to intermediate acceptability judgments, e.g., cyclic merge[37], late-cyclic merge?? and, post-cyclic merge. It is difficult to pursue this approach in more detail as there is no theoretical foundation that helps us to conceptualize what an intermediate timing of merge would be and how this would correspond to intermediate acceptability judgments of movement. Thus, this solution seems ad hoc and conceptually problematic. Instead maintaining a binary division between cyclic and post-cyclic merge such that complements and subjects are merged cyclically while adjuncts are merged post-cyclically, forces a stipulation that causal *fordi* 'because'-clauses are adjuncts, while conditional *om* 'if'-clauses are not.[38] Conditional *om* 'if'-clauses are not canonical complements as they are not selected by the verb. Consequently, it is not clear what phrasal category they should be assigned if they are assumed to neither be adjuncts nor complements. Thus, an explanation in terms of Late-Merge seems to raise more issues than it can resolve.

Truswell's (2007, 2011) assumption that all tensed adjuncts are islands due to a tensed operator blocking extraction also straightforwardly predicts that *fordi* 'because' is an island as *fordi* 'because' introduces a tensed clause. However, just like the Late-Merge approach, Truswell's (2007, 2011) assumption about tensed clauses makes the incorrect prediction that extraction from *om* 'if' should be illicit. Assuming that conditional *om* 'if'-clauses do not include a tense operator is not possible. Conditional *om* 'if'-clauses cannot be non-finite and tenseless. Both causal *fordi* 'because' and conditional *om* 'if' can occur in both the past and the present tense.

An additional difficulty that the categorical approaches face is the finding that extraction from *om* 'if' varies by type of A'-dependency. In the current thesis, I have shown that *rc*- and top-dependencies yield the same intermediate island effect with *om* 'if'. However, it has been shown that forming bare *wh*-dependencies into *om* 'if' yields classic island effects (Kush et al.

[37] Pre-cyclic merge is impossible as elements are thought to enter the derivation cyclically.

[38] There is ongoing debate as to whether or not there really is evidence to maintain a clear distinction between complements and adjuncts for phrase structure (see McInnery 2022 for an interesting discussion of this).

2018). It is impossible within the categorical approaches to explain this difference between A'-dependencies. It does not make sense to assume that *om* 'if' is cyclically merged in *rc*-dependencies, but late-merged in *wh*-dependencies as there is no determining characteristic that forces merge of the adjunct clause at an earlier stage in a *wh*-dependency compared to in an *rc*-dependency (at least not in current theorizing on A'-dependencies).

It seems very clear that categorical approaches like Stepanov's (2001, 2007) Late-Merge or Truswell's (2007, 2011) tensed clause assumption cannot explain the patterns uncovered in the current thesis. Overall, the empirical findings suggest that categorical approaches to extraction from adjunct clauses cannot be applied to explain the extraction patterns. However, there are some very clear generalizations that can be made from the data, which strongly indicate that there are categorical distinctions between factors, just not ones that rely on the insights of the CED:

- *Om* 'if' consistently yields acceptability judgments different from *fordi* 'because' and *når* 'when'.
- Top- and *rc*-dependencies are the same for extraction, while *wh*-dependencies are not.

This work shows that certain A'-dependencies cluster together to the exclusion of other A'-dependencies, and that there is variation between adjunct clause types for the same A'-dependencies. Thus, we need a theory that is sensitive to both these distinctions.

The non-categorical approaches allow us to pursue an explanation for the differences between adjunct clause types, while maintaining that adjunct clauses that partially allow extraction are adjuncts, as opposed to another phrasal category. The syntactic approaches assume that there are finer-grained syntactic distinctions between dependencies and clause types, while discourse-based approaches assume pragmatic distinctions between syntactically similar constructions. I think there is strong evidence to reject discourse-based approaches to my data, at least as Goldberg (2006) and Abeillé et al. (2020) present them. The prominent discourse-based approaches make incorrect predictions, indicating that a syntax-based non-categorical approach is favorable. For instance, Abeillé et al. (2020) predict that topicalized elements (i.e., focused) should not be part of (dependency-formation) backgrounded constituents, but that *rc*-dependencies (i.e., non-focused) should be permissible into constituents of all discourse statuses (see discussion in Section 3.2.3), the results for *fordi* 'because' and *når* 'when' directly contradict this prediction. In other words, that *fordi* 'because' and *når* 'when' are islands for relativization is not predicted under Abeillé et al.'s (2020) account. Kobzeva et al. (2022) also find that the FBC constraint makes incorrect predictions for

extraction with *wh*- and *rc*-dependencies in Norwegian. Furthermore, Nyvad, Müller & Christensen (2022) find that discourse-based approaches alone cannot provide an adequate explanation of their findings for adjunct extraction in English.[39]

An apparent strength of the discourse-based approaches is that they provide straightforward explanations of gradience in acceptability. The notion of backgroundedness is assumed to be gradient (Goldberg 2013; Namboodiripad, Bisnath, Kramer, Luntzlara & Goldberg 2021). Consequently, the extent to which clauses are islands should covary with the level of backgroundedness, both of which will be gradient notions. The intermediacy displayed in the ratings of *om* 'if' is highly consistent (something which is also seen in the raw ratings; see Section 4) between participants and items. Syntax-based approaches derive island effects as functions of more general properties of syntax, predicting uniformity between and within participants for sentences that are the same for the relevant factors. This is exactly what is seen in the current experiments. As discourse factors such as backgroundedness are not controlled for across items within the same adjunct clause type, I assume that the consistency across items is most compatible with a syntax-based approach, as opposed to a discourse-based approach.

The current non-categorical syntax-based approaches explain adjunct islands as a function of the external syntax of adjunct clauses, deriving differences between adjunct clauses as differences in the *height* of merge. These height-based approaches can capture differences between adjunct clauses. For instance, Haegeman (2012) shows that adjunct clauses that are merged very high in the matrix clause have a peripheral status and resist extraction, while adjuncts merged low in the structure have a central status relative to the matrix predicate and they can allow extraction (see also Müller 2019). This account predicts that there should be differences in the height of merge between *fordi* 'because' and *når* 'when' on the one hand, and *om* 'if' on the other. However, Paper 1 finds that there is no distinction between *fordi* 'because'- and *om* 'if'-clauses on the parameters that Haegeman (2012) provides to distinguish between central and peripheral adjunct clauses. Also, and more seriously, additional assumptions about how the height of merge and locality are related are required for this line of explanation to have any explanatory power.[40]

[39] Nyvad et al. (2022) also argue that it is difficult to operationalize discourse-based approaches in a falsifiable way.

[40] Boeckx (2012) assumes that adjuncts merged at the phase edge are Pair-Merged (i.e., opaque), while adjuncts merged below the phase edge are Set-Merged (i.e., transparent). It is difficult, however, to see how this approach can maintain a distinction between low adjuncts and complements (if such a distinction is needed). Also, this approach, when applied without additional constraints, predicts that extraction from *om* 'if' should be fully

5.2 A new syntax-based approach

The above discussion finds that none of the existing approaches to adjunct islands can explain the patterns revealed in the current project. Furthermore, closer scrutiny of the experimental results provides evidence that the constraint(s) responsible for the adjunct island effects uncovered in the dissertation is(are) syntactic in nature (due to their consistency across items and participants). I find in Paper 3 that the *internal syntax* of adjunct clauses is sufficient for explaining the difference in extraction patterns. This explanation relies entirely on two general and broad-ranging approaches to islands as opposed to adjunct-specific approaches. Thus, I find that existing theories *can* provide an explanation to the empirical patterns revealed, but that these are not adjunct-specific and that two traditional approaches must be applied in new ways to explain the empirical patterns revealed in the current thesis.

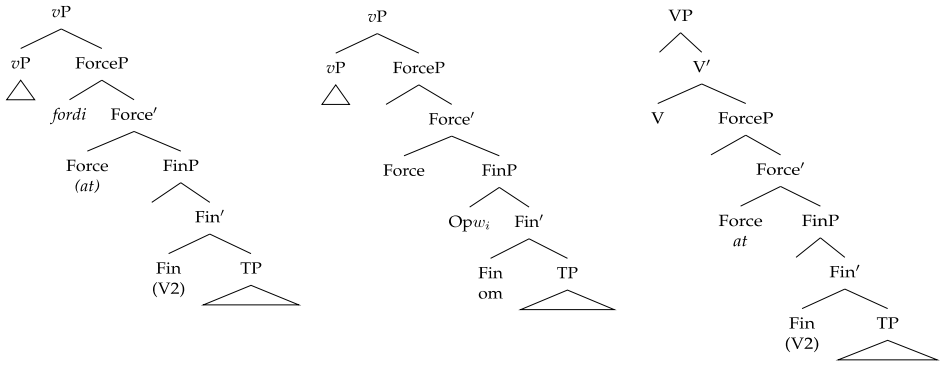
In Paper 3, I outline some details of the syntax of embedded clauses in Norwegian and A'-dependencies and adjunction in general. For details and evidence for the formal analyses of the adjunct clauses, please see Paper 3. I assume the following structure of embedded adjunct clauses in Norwegian, which rests on several general assumptions; Phase Theory (Chomsky 2000, 2001)[41] a cartographic analysis of the CP-domain (see Rizzi 1997), V2 asymmetry (see den Besten 1981; Holmberg 2015), a promotion analysis of relative clauses (Åfarli 1994), a movement derivation of conditional clauses (see Bhatt & Pancheva 2006; Haegeman 2010; Chatzopoulou 2021), and featural Relativized Minimality (fRM) (Rizzi 1990; Rizzi 1997; Starke 2001; Rizzi 2001).

acceptable, in the case that *om* 'if' is merged low, that is. I do not see how this approach can provide an adequate explanation of the current findings. However, just like Haegeman's (2012) approach, there might be other adjunct clause types for which these height-based approaches provide good explanations.

[41] Tough see discussion in Paper 3 that Phase Theory is simply the current implementation, general X'-theory and successive cyclic movement yield the same result on the current analysis. I do not favor Phase Theory in particular, but I have currently chosen this implementation as it is a mainstream implementation that provides the mechanisms that I find are required to explain the empirical findings.

Figure 7: Structural analysis of embedded adjunct clauses in Norwegian

A) *fordi* 'because' B) *om* 'if' C) *at* 'that'



First, the current dissertation finds that adjunction qua adjunction should have no consequence for extraction. Thus, it is expected that the way each of the clauses are merged with the matrix clause in Figure 7 should not matter for extraction.[42] Instead, following detailed evidence of word order and movement operations in the CP-domain of adjunct *fordi* 'because' and *om* 'if'-clauses, I argue that the differences in the internal syntax between the clauses in Figure 7 derives the differences in extraction patterns.

Second, as illustrated in Figure 7, I find that the internal syntax of causal *fordi* 'because' and conditional *om* 'if'-clauses differ in ways that matter for locality conditions. Based on evidence of embedded topicalization in *fordi* 'because'-clauses and evidence that *fordi* 'because' can precede the declarative complementizer *at* 'that' in *fordi at* 'because-that'-structures, I argue in Paper 3 that *fordi* 'because' is situated in Spec-ForceP. I assume Force to be a phase head in Norwegian (see also Klævik-Pettersen 2022), consequently blocking all extraction from *fordi* 'because'-clauses under the PIC. This analysis predicts that extraction from *fordi* 'because' by any A'-dependency type will be illicit in Norwegian. This fits well with the empirical results in the present thesis. For *om* 'if', I postulate, based on independent syntactic evidence, that there is an operator of possible worlds in Spec-FinP and that the complementizer *om* 'if' is situated in Fin. Consequently, there is no structural element that prevents movement to the phase edge (Spec-ForceP). Thus, the PIC does not constrain movement from *om* 'if'-clauses. On a representational level, however, the operator over possible worlds acts as a partial intervener between the two copies of the A'-moving elements

[42] One consequence of this could be argued to be that only one type of Merge is required. I will, however, leave this discussion for future work.

as it contains a [+Op] criterial feature. Partial overlap between the mover and the intervener, where the mover is more complex than the intervener (i.e., *rc*- and top-dependencies), is theorized to be acceptable (Starke 2001) and has been shown to yield acceptability ratings in the intermediate range in an experimental setting (Villata et al. 2016). This is exactly the pattern that is revealed in the current experiments for *om* 'if' (see Paper 3 for a detailed discussion of overlap patterns).

The current experiments provide evidence that *rc*- and top-dependencies pattern together to the exclusion of bare *wh*-dependencies. Kush et al. (2018) find that forming bare *wh*-dependencies into finite *om* 'if'-clauses yield classic island effects. This difference between dependency types follows naturally from the featural make-up of *wh*-dependencies. In bare *wh*-dependencies there is only a *wh*-word that moves, which will have a simple featural make-up. This yields full overlap between the *wh*-word and the possible worlds operator in *om* 'if'-clauses, and subsequently yield low acceptability ratings, just as observed in Kush et al. (2018).

Complex *wh*-dependencies, which have the same featural make-up as *rc*- and top-dependencies, should yield acceptability scores on a par with top- and *rc*-dependencies in *om* 'if'-clauses on the current account. However, Kush et al. (2018) find that they yield classic island effects with *om* 'if' in Norwegian. In Kobzeva et al. (2022), on the other hand, complex *wh*-dependencies yield small island effects in line with the predictions of the current proposal. One potential confounding factor in Kush et al. (2018) is that the *om* 'if'-clauses are tested with past tense. In all subsequent experiments with *om* 'if', the clauses tested are in present tense. As conditional *om* 'if' yields an irrealis reading, it might be easier to accommodate present tense with an irrealis reading than past tense. This could have an impact on Kush et al.'s (2018) results. Notice though, that this would perhaps also have an impact on the bare *wh*-dependencies, which are also tested with past tense. Accordingly, both *wh*-dependency types should be investigated more closely as there are possible confounding factors that influence the results. The current proposal predicts that bare and complex *wh*-dependencies should yield contrasting results, i.e., in line with Kobzeva et al.'s (2022) results for complex *wh*-dependencies.

The results for *når* 'when' are less clear than the results for the other two adjunct clause types. This adjunct clause type is not discussed in detail in Paper 3. *Når* 'when' yield variable ratings in a top-dependency. Variable ratings are not easily interpreted in the classic Sprouse-design as there is no way to distinguish between different sources of variation. Thus, we cannot know whether the variation seen in top-dependencies with *når* 'when' is caused by variation between speakers or items. I think *når* 'when' must be investigated more closely before any

conclusions on its island status can be made. However, I will sketch some preliminary ideas for how to interpret the results.

First, I assume that the difference in results between the top- and the *rc*-experiments is is caused by factors that are not meaningful for island phenomena (for instance caused by differences in naturalness of long sentences with topicalization vs. relativization). Given the similarity between the dependency types for *om* 'if', it seems most likely that the two dependency types should be the same for *når* 'when' as well.[43] There is no mechanism in the current proposal that distinguishes the two dependency types. They both involve movement to the embedded Spec-ForceP of a DP with a complex featural make-up. In other words, just like for *om* 'if', I assume that the two dependency types are the same for *når* 'when'.

Second, I assume that extraction from *når* 'when' generally yields classic island effects. The results for *når* 'when' in ExperimentRel2 are indistinguishable from the results for *fordi* 'because', and if there is no difference between *rc*- and top-dependencies, the results for *når* 'when' in *rc*-dependencies should extend to top-dependencies as well. Thus, in both top- and *rc*-dependencies, the main pattern for *når* 'when' is that of a classic island effect. Given the analysis provided for *fordi* 'because' and *om* 'if' above, this indicates that *når* 'when', just like *fordi* 'because', might be a specifier of ForceP blocking extraction. Interestingly, *når* 'when' can appear before *at* 'that' just like *fordi* 'because'.[44] The sentence pair in (62) shows that the interpretation of the sentence does not change when *at* is added.

[43] On the current proposal all dependency types, both featurally simple and complex ones, should yield the same results for *fordi* 'because'. Thus, restrictively, it is the evidence that top- and *rc*-dependencies are the same for adjunct *om* 'if'-clauses that provide evidence that these dependency types are the same.

[44] There is no evidence, however, that *når* 'when' can appear with V2 word order like *fordi* 'because' can. This might indicate that there are some differences between *fordi* 'because' and *når* 'when'. Whether these differences are relevant for islandhood remains to be investigated.

(62)

a. *Læreren* *blir* *sur når* *elevene* *jukser* *på prøven*
teacher.DEF becomes mad when students.DEF cheat on test.DEF
'The teacher gets mad when the students cheat on the test'

b. *Læreren* *blir* *sur når* *at* *elevene* *jukser* *på*
teacher.DEF becomes mad when that students.DEF cheat on
prøven.
test.DEF
'The teacher gets mad when the students cheat on the test'[45]

This provides some independent evidence that *når* 'when' could be a specifier of ForceP. Thus, *når* 'when' will be an island for all dependency types.

Still, there is considerable variation in the judgments for *når* 'when' that is not seen in *fordi* 'because' in the top-experiments. Under the analysis that *når* 'when' derivationally blocks all extraction, variation is not expected and must be explained. If we see differences between participants, one potential answer might be that participants analyze *når* 'when' at differently. Perhaps *når* 'when' is a specifier of ForceP for some participants, but e.g. a complementizer like *om* 'if' for other participants. If there are differences between items, one potential answer is that on trials where extraction is perceived as acceptable, there is a difference in the interpretation of *når* 'when' which has implications for the internal syntax. There might be small interpretative differences in the items that are not controlled for in the design.

In summary, the approach that I propose here is that adjunction qua adjunction does not matter for extractability and that two locality conditions are at work in complements and

[45] *Om* 'if' can also appear before *at* 'that' in Norwegian, but this is very rare. However, sentences with *om at* 'if that' do not have a conditional reading, which indicates that *om* 'if' in *om at* 'if that' is different from conditional *om* 'if' and should not have the same internal structure. I understand *om at* 'if that' in (iii) below as introducing a purely causal reading.

(iii) *om at* 'if that' does not have a conditional reading
Jeg *er* *glad* *om* *at* *dette* *kommer* *akkurat* *nå*
I am glad if that this comes exactly now
'I am happy that this comes right now'
(retrieved from: https://www.ntnu.no/universitetsavisa/0900/mjos.html)

adjuncts alike: there is an online, derivational locality condition *and* a filter on the output of the derivation (representational locality condition). Each of these conditions belong to the canonical syntax-based literature on islands and have been around for a long time. The new proposition in the current dissertation is that these conditions must be combined to adequately explain the empirical landscape. Traditionally, these conditions have been considered opponents and different locality effects have been attempted combined under one or the other locality condition (see the discussion in Epstein & Seely 2002; Boeckx 2012; Torr 2012 etc.). I instead propose that both locality conditions must work in tandem in order to explain locality effects. I will refer to the approach that I here suggest as the *The Double Trouble Locality Condition (DTLC):*

(63) *The Double Trouble Locality Condition*
(i) movement is successive cyclic targeting the edge of certain domains (e.g., phases). If a movement step is too long (e.g., forced to be too long due to an already filled position), the derivation will crash.
(ii) In any derivation, the filler and the gap must be uniquely identified. If there is featural overlap between two potential fillers, the representation will be perceived as (much) less acceptable.

(i) implemented as Phase Theory specifies that any A'-moving element must target the Spec of the phase edge. If the Spec of the phase edge is already filled, the A'-moving element cannot escape the phase and the derivation will crash. This locality condition is a derivational locality condition that rules out extraction from *fordi* 'because'-clauses across dependency types. (ii) implemented as fRM specifies that full overlap between the mover and the intervener yields unacceptability, while partial overlap yields somewhat acceptable *representations*. (ii) is a representational locality condition. (i) does not rule out extraction from *om* 'if' in Norwegian, but (ii) is responsible for degrading acceptability of extraction from *om* 'if'. For *når* 'when' we need more research, but it seems that (i) generally excludes extraction from *når* 'when'.

5.3 Theoretical implications

There are two main theoretical implications that follow from the current discussion. First, in order to explain the different island effects that the three adjunct clause types yield, both a derivational and a representational condition on locality are required. Second, even if the empirical investigation proves that a fine-grained theory is needed to explain patterns of adjunct

clause extraction, two very general locality conditions on the *internal* syntax of clauses provide an adequate explanation of the patterns.

Starting with the first point, the two locality conditions act in succession. The derivational locality condition determines which operations are possible in the derivational space. This constraint applies at the level of structure-building by providing certain rules for how structure-building can proceed. The outcome of a derivational constraint can either be successful – the structure is built, or unsuccessful – the structure is not built. This is compatible with a binary grammar. The derivational locality condition, implemented as Phase Theory in the current thesis, allows an element to escape from an *om* 'if'-clause, but not from a *fordi* 'because'-clause. A representational locality condition, which is here implemented as fRM, provides a filter on the structure that is derived.[46] The representational locality condition is only

[46] It is not clear at exactly what "time" a representational locality condition like fRM applies. There is nothing particularly phonological or interpretational about the condition, meaning that it is not immediately clear that fRM either is a PF or an LF phenomenon. Thus, perhaps the condition applies at Spell Out. There are many intertwined and complicated questions that arise here. For one, assuming that fRM acts as a constraint on chain-formation, the present story indicates that chain-formation occurs at Spell Out (see Epstein & Seely 2002a: 6-7 for a discussion of whether chains are necessary syntactic objects or whether they unnecessarily complicate the theory). If Spell Out is cyclic in the sense of Phase Theory, it follows that intervention effects can only occur if there are two featurally similar elements within the same phase. Given that movement is successive cyclic targeting each phase head, this might always be the case and so cyclic Spell Out might not be a problem for the present implementation of a representational filter in the form of fRM:

(iv) The Spell Out of phases

Phase 1: [$_{\text{ForceP}}$ X [$_{\text{Force'}}$ [$_{\text{FinP}}$ … [$_{\text{DP}}$ Z]]]]

Phase 2: [$_{\text{ForceP}}$ X [$_{\text{Force'}}$ [$_{\text{FinP}}$ Y [$_{\text{Fin'}}$ *om* 'if' [$_{\text{TP}}$ … [$_{\text{ForceP}}$ Z]]]]]]

Trying to form a dependency between X (filler) and Z (gap) is possible in Phase 1. Here no representational locality condition is violated. In the second phase, forming a dependency between X and Z is interrupted by Y. As the head of Phase 1 is still visible for the next phase (Phase 2), there are two competing elements that can enter the chain. This makes the derivation less computationally efficient. Epstein & Seely (2002b) propose that Spell Out is not intrinsically linked to phases, but instead that each "transformational" rule (e.g., application of Merge) activates Spell Out. If this means that only once X is merged in matrix Spec-ForceP will the features relevant for these positions be spelled-out, i.e., the representational filter will only apply once the full derivation is complete, the same result will, as far as I can see, be obtained.

(v) Spell Out once transformational operation has applied:

[$_{\text{ForceP}}$ X [$_{\text{Force'}}$ [$_{\text{FinP}}$ Y [$_{\text{TP}}$ … [$_{\text{ForceP}}$ Z [$_{\text{Force'}}$ … [$_{\text{DP}}$]]]]]]]

partially fulfilled in the case of extraction from *om* 'if' as there is partial overlap between the possible worlds operator and the A'-moving element in top- and *rc*-dependencies.

In the context of islands, it is interesting to see that the derivational locality condition will exclude all cases of extraction from structures in which there is a filled phase head specifier (e.g., Spec-ForceP), while the representational locality condition can apply to cases where there is extraction from a structure with an open escape-hatch. In that sense, neither of the conditions are redundant, but they have complementary jurisdictions. Traditionally, the problem that the Subjacency Condition, *Barriers* and the PIC have been faced with, is cases of degraded extraction where there is no derivational reason why extraction should be illicit, e.g., adjunct islands (Subjacency Condition, PIC). Similarly, representational locality conditions have been challenged by cases where extraction is illicit, but there is no obvious intervening element, e.g., adjunct *fordi* 'because'-clauses. Furthermore, the derivational approaches predict a binary division between islands and non-islands, but the empirical realities do not comply with this system. Incorporating a non-binary system within the purely derivational accounts has proven to be difficult, and as such, it seems like a better solution to allow two different locality conditions to apply together – one derivational and binary, the other representational and gradient. This system should be strong enough to exclude all cases of illicit extraction that are syntactic in nature, while still allowing cases of licit and partially licit extraction. In addition, it should be flexible enough to apply to different types of clauses differently and to make fine-grained predictions (i.e., predicting differences between dependency types).

Lasnik (2001) & Boeckx & Lasnik (2006) also argue that both a derivational and a representational locality condition are required to adequately explain island phenomena. Their argument is based on cases of island repair. Island repair are cases where illicit extraction has taken place, but where the deletion of the offending trace (e.g., sluicing) (Lasnik 2001; Boeckx & Lasnik 2006) or insertion of a dummy-pronoun (resumptive pronoun) or dummy-gap (parasitic gap) (Engdahl 1983) repairs the sentence such that it is perceived as acceptable. Boeckx & Lasnik (2006) argue that cases that cannot be repaired by sluicing are derivational islands, while cases that can be repaired by other mechanisms are representational islands. As such, Boeckx (2008a) suggests that it might be false to unify all islands under one umbrella, and that instead, derivational islands are fundamentally different from representational islands.

This is just to illustrate that there are many ways in which the general proposal that a representational locality condition applies at Spell Out can be implemented.

Many authors, however, argue that postulating two locality conditions raises a redundancy issue. This is an important issue, especially within the MP. Epstein & Seely (2002b) argue that there is only a derivational locality condition, and that what seems to be representational locality effects are really derivational. One of the central arguments in favor of the DTLC is that there is strong evidence in favor of a gradient condition. In my opinion, there is no way within a strictly derivational account to explain gradient results. The opposite view is argued for in Brody (2002), Boeckx (2012) and Torr (2012). These authors argue instead that there is only a representational locality condition at work, and that this representational locality condition must be more general than fRM.[47] The representational condition sketched in Boeckx (2012) is one in which all adjuncts and subjects will be islands, as all cases of adjunction induce immediate Spell Out. This system makes incorrect predictions given the empirical landscape of both subject and adjunct extraction. I think that an important reason why I can afford to postulate that there are two locality conditions at work, is that the derivational locality condition that I assume is very general. The only mechanisms that the current proposal requires from the derivational locality condition are the basic building blocks that movement is successive cyclic *and* that merge of two elements at the same node is illicit.

Postulating that there are two separate locality conditions at work also entails that there are (at least) two types of islands. The current thesis shows that causal *fordi* 'because'- and conditional *om* 'if'-clauses are two separate island types, a derivational island and a representational island, respectively. There is some initial evidence that the different types might have different properties beyond those discussed in this dissertation. Chaves & Putnam (2020), investigating English, find that judgments of extraction from *if*-adjuncts satiate after repeated exposure in an acceptability judgment experiment, while judgments of extraction from *because*-adjuncts do not. This might indicate that different types of islands have different properties.

[47] Torr (2012) for instance, implements a representational locality condition in a way reminiscent of a derivational locality condition. He uses metaphors like "make the move illicit" (2012: 104) for example. To me, it is problematic to describe a representational locality condition is such terms, as this indicates that fRM for instance has the strength to disallow movement. Under such an implementation of fRM, it is not clear to me how fRM can describe gradient outcomes. Is the move across an operator over possible worlds, like what we see in *om* 'if'-clauses, partially licit? In what way is it somewhat possible to move and how is such partially licit movement encoded? I think such an implementation of a representational-locality-condition-only view is weaker than a "mixed model" (Brody 2002: 19) where a derivational locality condition conditions structure-building, and a representational locality condition provides a filter on structure-building.

The proposal outlined above also indicates that adjunct clauses can be constrained by the same conditions that other clauses (i.e., complex NPs, *wh*-clauses etc.) are. In other words, the main theoretical findings of the dissertation provide evidence that, at least for the phenomena currently explored, making reference to the notion of "adjunct" is not necessary to explain the extraction patterns. Whether nor not this generalization can be applied to all cases of extraction from adjuncts remains to be seen, but if it can, it would mean that there is nothing intrinsically "island-y" about adjuncts. This would be in direct contrast to previous work on the AC following Huang (1982) as well as the non-categorical height-based approaches (Haegeman 2012; Brown 2015a,b; Müller 2019). The current proposal illustrates that the internal syntax of adjunct clauses is sufficient for deriving the island effects displayed by both *fordi* 'because' and *om* 'if'. This means that the current dissertation raises questions about notions that adjuncts are merged later or in a different way in the sense that such mechanisms automatically make the adjunct impossible to extract from. That said, it does not mean that the category of adjuncts can be disposed of. There are still other important distinctions between complements and adjuncts, but one of the defining differences between the two phrasal categories, namely that one is an island and the other is not (see e.g., Bode 2020: 17), can be disposed of.

Another and highly important caveat to the present work is that not all cases of unacceptable extraction from adjuncts must be explained syntactically. The DTLC dictates that there are two syntactic conditions working in tandem, but does not exclude the possibility that there are additional semantic/pragmatic and/or phonological conditions that might also interfer with acceptability. The empirical evidence that the present thesis relies on consists of highly similar test sentences that are matched on several syntactic and semantic factors. Thus, that there might be cases of *illicit* topicalization from *om* 'if'-clauses is not necessarily counterevidence to the present proposal, but might instead suggest that there are additional (extra-syntactic) conditions that apply. For instance, finding (for example) that extraction is impossible from conditional *om* 'if'-clauses that are in the past tense might suggest that there is a semantic filter on tensed adjunct clauses akin to Truswell's (2007, 2011) semantic condition on untensed adjunct clauses, which applies in addition to the syntactic conditions. However, finding examples of licit topicalization from *fordi* 'because'-clauses *in Norwegian*, without any evidence of grammaticality/acceptability illusion at play, would be highly problematic for the present proposal. This follows from the current proposal as extraction from *fordi* 'because'-clauses should be ruled out by a derivational constraint, and therefore semantic, pragmatic and/or phonological factors should not matter for acceptability. Finding examples of licit extraction from *fordi* 'because'-clauses in other languages is not direct counterevidence for the

present proposal unless there are clear indications that the internal syntax of causal clauses is the same across languages.

5.4 A comparative view

That finite adjuncts are not always islands and that all A'-dependencies do not pattern in the same way are empirical findings that are not only true for Norwegian. The empirical evidence has revealed that, in most languages investigated, adjunct island violations are allowed under certain conditions. Specifically, it has been shown that dependency types might differ in their island sensitivity cross-linguistically. A general pattern that has emerged is that finite adjunct clauses are islands in a *wh*-dependency (Kohrt et al. 2020, Sprouse et al. 2016, Kush et al. 2018, Sprouse et al. 2012a), but might not be so in an *rc*-dependency in English (Sprouse et al. 2016; Nyvad et al. 2022) or Norwegian (Paper 2) or in a top-dependency in Norwegian (Kush et al. 2019, Paper 1), Danish (Müller & Eggers 2022), Swedish (Müller 2019) or Chinese (Myers 2012; Zenker & Schwartz 2017). Furthermore, that adjunct clause types are not the same for extraction seems to also be a pattern that is found across languages (Müller 2019; Paper 1; Nyvad et al. 2022; Müller & Eggers 2022; Paper 2). The current section will investigate and discuss how the empirical and theoretical findings of the dissertation compare cross-linguistically.

First, the DTLC is predicted to be cross-linguistically valid as it builds on general properties of language. Second, the DTLC predicts that there will be differences in the island inventory between languages, given that there might be differences between languages regarding central features such as the place of merge for complementizers etc. Furthermore, if it can be shown that A'-dependencies have a different featural make-up cross-linguistically, the DTLC also predicts that there will be differences in how A'-dependencies impact islandhood. Subsequently, the current proposal predicts that the same conditions will apply cross-linguistically, but, as the conditions are as flexible as the internal syntax of embedded clauses is across languages, differences between languages can occur. Below I investigate some empirical findings in other languages to see how the DTLC fares with cross-linguistic adjunct clause extraction patterns.

A very similar study to the one reported in Paper 2 is carried out for English by Nyvad et al. (2022). Nyvad et al. (2022) test adjunct clauses introduced by *if*, *when*, and *because* in a demonstrative *rc*-dependency, i.e., introduced by *this is*. They find the same results for English that I find for Norwegian: *because* and *when* pattern together to the exclusion of *if.* For *because* and *when* they find low acceptability, and much lower acceptability than for each of the control conditions (no extraction + adjunct clause, extraction + declarative clause, no extraction +

declarative clause). *If* yields acceptability scores on a par with extraction from declarative *that*-clauses. Müller & Eggers (2022) find the exact same patterns for Danish and English in a corpus study. For Danish they find examples of topicalization and relativization out of conditional adjunct clauses introduced by *hvis* 'if'. They find no examples of *wh*-movement out of *hvis* 'if'. Also, they find no examples of extraction from *fordi* 'because'-clauses in the Danish corpus. For English they only find cases of relativization out of conditional *if*-clauses, no cases of *wh*-movement or cases of extraction (of any kind) from *because*-clauses. Interestingly, for both languages they find a very small number of examples of extraction by way of relativization (English and Danish) and topicalization (Danish) from *when*-clauses.

The proposal developed to explain Norwegian in the current thesis can be applied to Nyvad et al.'s (2022) and Müller & Eggers (2022) findings for English and Danish. I assume the exact same internal syntax of the adjunct clauses in Danish as in Norwegian. Similarly, for English too we can propose that *because* and *when* occupy Spec-ForceP and derivationally block extraction. Historically, *because* could be followed by *that* or *why* in English (OED "because"). This indicates that *because* might be placed high in the structure, just like *fordi* 'because' in Norwegian. The current proposal predicts that extraction by any A'-dependency type should be derivationally illicit from both *because*- and *when*-clauses. This is what Nyvad et al. (2022) and Müller & Eggers (2022) find for English and Danish.

If-clauses, on the other hand, are best analyzed as being derived by operator movement to Spec-FinP, just like in Norwegian. The majority of evidence in favor of a movement analysis for conditional clauses comes from English (see Bhatt & Pancheva 2006). This analysis predicts that all types of featurally complex dependencies (such as *rc*-dependencies) into *if*-clauses should yield small island effects. This is the pattern that I consistently see in my experiments with top- and *rc*-dependencies, and the pattern found for English in a demonstrative *rc*-dependency in Nyvad et al. (2022). Similarly, Sprouse et al. (2016) find evidence that conditional *if*-clauses in an *rc*-dependency do not yield island effects. This indicates, within the logic of the Sprouse-design, that *if*-clauses are not grammatical islands in English. However, Sprouse et al. (2016) find very low average acceptability judgments for the island violating condition, unlike what I find in Paper 2 and what Nyvad et al. (2022) report for English. Thus, the studies find quite different judgment patterns for the same adjunct clause type. This illustrates that we need more research on this adjunct clause type in an *rc*-dependency in English, as well as more research into potential differences between types of *rc*-dependencies. Even if there are differences in the absolute judgment of the island violating condition, the

various experimental results minimally show that there is no evidence for a derivational locality condition constraining movement from conditional *if*-clauses in English.

Sprouse et al.'s (2016) results for conditional clauses in an *rc*-dependency in Italian do not match the pattern revealed for *rc*-dependencies into conditional clauses in Norwegian, English, or Danish. Sprouse et al. (2016) find that unlike *if*-clauses in English, *se* 'if'-clauses in Italian yield classic island effects in an *rc*-dependency. On the current proposal, this is surprising. Within the proposed DTLC, there are two main ways that Italian could differ such that the observed extraction patterns would be predicted; (i) that conditional clauses and A'-dependencies have different featural make-up cross-linguistically or, (ii) the internal syntax of the conditional clause differs cross-linguistically. Haegeman (2010: 599) suggests that the operator of possible worlds in conditional clauses moves to a higher position in the left periphery (specifically Spec-CP, my Spec-ForceP), as opposed to staying in Spec-FinP as I have suggested in the current proposal. This indicates that the operator can be situated in a high position in the left-periphery. There is also evidence that topicalization inside the conditional clause is possible in Italian (Munaro 2005) such that the topicalized elements directly follow *se* 'if':

(64) Topicalization in Italian

Se	***queste***	***cose***	*non*	*le*	*sai,*	*non*	*supererai*
If	these	things	not	CL	know.PRES.2SG	not	pass.FUT.2SG

l'esame.
the exam
'If these things you don't know, you won't pass the exam.' (Munaro 2005: 83)

Given the cartography of the left-periphery as established in Rizzi (1997), the internal topic position is situated between Force and Fin. Subsequently, if topicalization to a position just below the complementizer is possible in Italian conditionals, there is evidence that the complementizer is situated in Force and presumably the operator of possible worlds will be attracted to Spec-ForceP. Thus, if the operator moves to Spec-ForceP in Italian, the derivational locality condition will rule out extraction from conditional adjunct clauses as the escape hatch is occupied by the operator. I think that this could be a possible explanation, but for the account

to be well founded, independent evidence that conditional clauses are derived in this manner in Italian is required.[48]

There is also evidence that Czech allows extraction from conditional clauses on a par with extraction from non-finite purpose clauses, the latter constituting a classic example of non-finite adjunct clauses allowing extraction in English (Truswell 2011). Biskup & Šimík (2019) show that extraction from left-adjoined conditionals is allowed in Czech. This extraction pattern can be explained by the current thesis if conditional clauses in Czech have the same internal syntax that I assume for conditional *om* 'if'-clauses in Norwegian. The investigation in Biskup & Šimík (2019) is not experimental and whether extraction from conditional clauses in Czech also is slightly degraded as in Norwegian is not addressed. The pattern in Czech might be somewhat more complicated than in Norwegian. Biskup & Šimík (2019) show that extraction is only allowed out of left-adjoined adjuncts as opposed to integrated adjuncts (similar patterns are also found for English, see e.g., Boeckx 2012: 68). This extraction pattern is seen across different clause types and is not unique to conditionals. This seems to be an interesting venue for further investigation into the findings in Czech for existing height-based approaches to adjunct clauses as it shows that adjunction site matters for extraction.

Korean is a language that does not show any adjunct island effects. Such a pattern could be an effect of *wh*-words remaining in situ in Korean. However, since Huang (1982) covert movement of the A'-moving element at LF in *wh*-in-situ languages is typically assumed. Part of the evidence for this is based on findings that there are island effects in *wh*-in-situ languages (Huang 1982). For instance, Kim & Goodall (2016) show that there are consistent *wh*-island effects in Korean. Thus, that Korean does not display adjunct island effects cannot be explained as an effect of the *wh*-word remaining in situ. An analysis in terms of the DTLC will hinge on

[48] That there are island relevant syntactic differences between Norwegian and English on the one hand and Italian on the other is also supported by findings that complex subjects yield opposite results across the same languages. Complex subjects do not yield island effects in an *rc*-dependency in Italian, while they yield classic island effects in an rc-dependency both in Norwegian (Kobzeva et al. 2022; Paper 2) and in English (Sprouse et al. 2016). It is interesting and potentially theoretically relevant that the languages show opposite behavior on two island types that were thought to be derived by the same condition under the CED (Huang 1982). I do not currently have any ideas as to how this distinction between languages is relevant, but it suggests that there might be derivational and/or representational differences between languages with the result that the inventory of island types differs between languages. As Sprouse et al. (2016) find that subject islands are the same for English and Italian in *wh*-dependencies, it might be the case that the difference between languages regarding the subject island is featural. On the DTLC, a derivational constraint is not hypothesized to vary between dependencies, while a representational constraint is.

the analysis of *wh*-in-situ. If we assume that *wh*-in-situ involves movement at LF, there should be no difference in derivational locality constraints between languages with *wh*-in-situ and *wh*-in-C – if a node is already filled, both covert and overt movement should be ruled out. Looking at the representational filter is somewhat more complicated. Attempting to apply the DTLC to the Korean data raises questions about the timing of the representational filter. If fRM applies at Spell Out, and the *wh*-element moves at LF, a representational locality condition will not apply in *wh*-in-situ languages. If we assume, however, that fRM instead applies at LF, the pattern for *wh*-islands in Korean follows from the DTLC. As there are two *wh*-words with matching featural make-up in sentences with *wh*-island violations in Korean, these structures are degraded under the representational locality condition. In the adjunct clauses, however, there is no overlap (or only partial overlap in the case of conditional clauses) between features. In fact, the current account would predict that extraction from causal *because*-clauses should be more acceptable than extraction from conditional *if*-clauses as causal *because*-clauses are not derived by operator movement. If we instead assume that *wh*-in-situ is the same as *wh*-in-C, only that the high copy is deleted as opposed to the low copy, we can assume that movement occurs at narrow syntax and that *wh*-in-situ languages should essentially be the same as *wh*-in-C-languages. This seems to be the analysis that Chomsky (2000, 2001) assumes for covert movement, i.e., no independent cycle for covert movement, but deletion of one of the copies in the chain at Spell Out. Under this analysis of *wh*-in-situ, that there are no island effects of extracting from adjunct clauses in Korean is unexpected on DTLC. It is expected that Korean should yield the same pattern as *wh*-in-C languages such as Norwegian, so long as the internal syntax of each of the clauses in question is the same across languages.

Chinese is another *wh*-in-situ language that has been claimed to show insensitivity to adjunct island effects. Zenker & Schwartz (2017) for instance find that topicalization out of *because*-adjuncts is possible in Chinese. An explanation for this lack of an effect depends on general facts of Chinese A'-dependencies. For topicalization there is discussion whether there is overt movement in Chinese (see Myers 2012 for argument that topicalization involves overt movement and Liejiong 1990 that it does not). If topicalization involves overt movement, a derivational locality constraint would be operative. Thus, that *because*-clauses allow topicalization is an indication that the internal syntax of *because*-clauses is different in Chinese compared to in Norwegian (and English). However, as already stressed, such an analysis must be accompanied by independent evidence of the internal syntax of *because*-clauses in Chinese. If such evidence shows that *because* might be a specifier of the phase head in Chinese as well, this provides counterevidence to the current proposal. If topicalization does not involve overt

movement, however, that extraction from *because*-clauses is licit follows straight-forwardly: The representational locality condition is the only one that can apply and there is no evidence that *because*-clauses are derived by operator movement, thus, there is nothing that can constrain topicalization from *because*-clauses.

The above discussion shows that the DTLC allows flexibility in the island inventory across languages. Furthermore, it shows that the condition relies on island-independent facts about A'-dependencies and the internal syntax of embedded clauses in each language in question. Thus, the approach is not categorical insofar as it does not assume the same adjuncts to be islands cross-linguistically. Moreover, the comparative investigation reveals that there are still several details of the DTLC that remain to be explained and explored. For instance, the question of the timing of the representational locality condition is relevant for how the condition can explain island effects in *wh*-in-situ languages.

6 Future research

This thesis investigates the extent of fine-grained variation in adjunct island effects in Norwegian, looking at different adjunct clause types and dependency types. The findings indicate that we need a theory of adjunct islands that is sensitive both to differences between adjunct clause types and dependency types. A secondary goal of the thesis is to provide an explanation of the fine-grained variation. I propose that the internal syntax of each adjunct clause type explains the extraction patterns, and that interactions between the adjunct clause types' internal syntax and the syntax of different dependency types also follow naturally from the current proposal. Consequently, I argue that "adjunction does not an island make". More so, the thesis shows that both a derivational and a representational constraint on filler-gap dependencies are required: certain adjunct clauses are derivational islands yielding classic island effects, while others may be representational island effects yielding island effects ranging from classic island effects to small island effects depending on the featural make-up of the adjunct clause and the dependency type.

The main findings in this thesis open a new horizon for investigation. First and foremost, it shows that carrying out rigorous experimental investigations to assess and establish extraction patterns is highly beneficial as it can uncover patterns that had previously not been discussed in detail. In addition, it provides very strong evidence that adjunct clause types differ in extraction patterns, and that consequently, formal experimental work must be carried out on all adjunct clause types that are of interest, as well as with all dependency types of interest. The theoretical findings of the current project provide clear predictions that can guide future experimental investigation. However, I wish to explicitly say that the current proposal is developed as an explanation to the robust difference between conditional *om* 'if'- and causal *fordi* 'because'-clauses in Norwegian *only*. This is a very limited empirical basis, which is both a strength and a weakness of the current proposal. It is a strength because it provides an explanation to a distinction that is shown repeatedly to be relevant in the grammar. As the empirical base is very limited, this distinction is investigated in detail and the robustness of the findings provides convincing evidence that this is a distinction in acceptability that really should be mirrored in the grammar. However, although the proposal very easily can be extended to other adjunct clause type and even other island types, the proposal is not developed with this intent. I make some high-level claims that we need both a derivational and a representational locality condition, but this is again based solely on my strictly defined empirical scope. Therefore, the current proposal makes predictions that should be rigorously investigated in future work but I do not claim that all island effects can be explained by the current proposal.

Here are some obvious predictions made by the current proposal that are worth making explicit and which can be investigated in an attempt to see the general applicability of the current proposal: (i) *rc*- and top-dependencies will be the same for all clause types; (ii) adjunct clauses derived by operator movement to Spec-FinP on a par with conditional clauses will yield intermediate ratings in featurally complex dependencies and classic island effects in featurally simple dependencies; (iii) adjunct clauses where there is a phrasal element in the specifier of the phase head will yield classic island effects; (iv) there will be variation in adjunct clause extraction patterns between languages in which the adjuncts in question have diverging internal syntax. This prediction is discussed to some extent in Section 5.4 above. These are predictions made by the current proposal that should be tested in future work.

The current proposal also predicts that there should be no difference in extraction patterns between complement and adjunct clauses which have equivalent internal structures. Going outside the limited scope of adjunct clauses, subject- and *whether*-clauses were included in the four of the five experiments carried out for this project as control clauses. I will here provide a short discussion of how the DTLC extends to these clause types. Starting with complement *om* 'whether'-clauses, in each experiment, the *om* 'whether'-clauses yield smaller island effects than adjunct conditional *om* 'if'-clauses. If embedded interrogative clauses introduced by *om* 'whether' are derived by the same mechanisms as adjunct conditional clauses, we would expect these clause types to yield the same results. There are two lines of inquiry here which should be pursued. First, minor differences in acceptability scores should be handled carefully. We do not have good models for interpreting acceptability ratings and in particular what minor differences in acceptability reflect. Second, Bhatt & Pancheva (2007) assume that only *adjunct* interrogative clauses have a conditional interpretation and that *complement* interrogative clauses do not. Thus, we might predict that complement *om* 'whether'-clauses are not derived in the exact same way as conditional *om* 'if'-clauses, and they might have different internal syntax compared to conditional *om* 'if'-clauses. Therefore, complement *om* 'whether'- and adjunct *om* 'if'-clauses should be tested in an experimental design set up specifically to compare these two clause types, and perhaps with a behavioral measure other than acceptability judgments on a 1-7 scale, e.g., forced choice task or perhaps an online measure such as self-paced reading etc.

Complex subjects, on the other hand, yield classic island effects across all dependency types tested in Norwegian – bare *wh*- (Kush et al. 2018), complex *wh*- (Kush et al. 2018), top- (Kush et al. 2019; Paper 1) and *rc*-dependencies (Paper 2). Stepanov (2007) already disconfirms that extraction from subjects should be ruled out under the CED. He shows that there are

languages that do allow extraction from subjects. Norwegian seems to not be such a language as the experimental evidence consistently shows that extraction from subjects yield classic island effects. It is not obvious, however, how subject islands can be explained by the general locality conditions outlined in the current proposal. It is not immediately clear that there is a filled Spec-ForceP or that there are any intervening elements that fully match the features of the mover. Lohndal (2009) shows that extraction *of* the subject is to some extent permitted in Norwegian, even if there is a complementizer immediately preceding the subject, so the position in itself might not be subject to freezing effects in Norwegian. It might be that complex subjects, typically having a topic status in the discourse, have a complex feature make-up, just like other NPs that are topicalized, relativized etc. This would mean that extraction of an NP from inside the subject clause would yield full overlap of features with the edge of the subject clause, yielding full overlap of features and a representational violation. Such an account predicts that in languages where extraction from a subject is permitted, the subject (intervener) is not as featurally complex as the dependency. This is a very stipulative suggestion and should only be read as an illustration of how the current proposal could be applied to other island types.[49]

Finally, continuing this line of work requires that we reach new understandings of the conceptual foundation of a representational locality condition like the fRM. In Minimalism, only conditions which have conceptual motivation should be included in the theory. The conceptual motivation for Phase Theory is that the lexicon is too demanding (Chomsky 2000: 100-101) to keep active during the whole derivation, and therefore, only small parts of the derivation is built at once from lexical sub-arrays (Chomsky 2000: 106, Boeckx & Grohmann 2007: 205). This means that the lexicon only needs to be activated during short intervals, and not during derivational operations such as movement etc. As pointed out in Boeckx (2012), a

[49] For instance, Italian has been shown to allow extraction from subject clauses only in an *rc*-dependency (featurally complex A'-dependency) and not in a simple *wh*-dependency (Sprouse et al. 2016). The same is true for French (Abeillé et al. 2020). As there is a difference in dependency type, the current proposal predicts that this difference must be representationally conditioned. The featurally complex dependency type is allowed into complex subjects. Thus, the present proposal predicts that complex subjects must have simple featural make-up in Italian and French that creates a pattern of full overlap with simple *wh*-dependencies. I have not found any evidence in support of such a representatioinal analysis for Italian and French, however. For one, Rizzi & Bocci (2017, fn. 9) show that pre-verbal subjects are not the same as topics in Italian. This could indicate that pre-verbal subjects do not have a [+Op] feature in Italian, which would mean that a full-overlap pattern with simple *wh*-dependencies cannot be the cause of the low acceptability. I think there are many interesting points that must be investigated here which could shed light on how fRM works and how complex subjects are and are not constrained cross-linguistically.

representational locality condition like the fRM comes close to the Minimalist core principles of efficiency, non-redundancy, least effort and the like. However, it is not clear what part of the Language Faculty that is aided by a correct representation. For instance, can relativized minimality effects be reduced to processing effects? There are many interesting venues for future research in establishing exactly how a representational filter works and how it (cognitively) aids language derivation.

7 Concluding remarks

Papers 1, 2 and 3 together with the cover article make up this dissertation. Papers 1 and 2 are empirically directed, while Paper 3 is theoretically directed. The papers are closely linked by first seeking descriptive adequacy of extraction patterns for certain adjunct clauses in Norwegian and secondly seeking explanatory adequacy of the empirical patterns uncovered in the first papers. Papers 1 and 2 follow closely in the footsteps of previous work within psycholinguistic/experimental syntactic research. The tradition here is to discuss the empirical findings within existing (syntactic) approaches (see e.g., Sprouse et al. 2016 or Kush et al. 2018, 2019). Paper 3 seeks to unite the experimental syntactic work with theoretical syntactic work by limiting the scope of data to experimentally collected data and trying to adapt and extend existing theories to fit the new empirical discoveries. As such, the paper finds itself at the crossroads between experimental and theoretical syntax. An auxiliary goal of Paper 3 is to combine these approaches without compromising the quality, principles, and methodology of either.

The papers comprising this thesis, alongside the cover article, contribute to several important discussions in the field. First and foremost, Papers 1 and 2 provide robust evidence that there is fine-grained variation in extraction patterns that must be accounted for. Although the papers are limited to investigating Norwegian, the findings extend to the general theory of adjuncts within the generative framework. If Norwegian is sensitive to these distinctions, we need a theory that can accommodate such distinctions, though not necessarily for all languages (i.e., parametric variation). Thus, the findings indicate that theories of islands need to be more fine-grained than traditionally proposed. As such, the thesis aligns nicely with current trends in island research where more fine-grained theories are being proposed to suit more nuanced empirical data. There is general debate as to what characteristic of linguistic expression that govern extraction patterns – are island constraints syntactic, semantic, or pragmatic in nature? The thesis points in the direction that constraints on finite adjunct clauses first and foremost are syntactic. Furthermore, there is a long-standing tradition treating adjuncts and complements as antipodes, but the papers in the current thesis provide arguments that there is not an inherent distinction between constraints governing adjunct clauses and complement clauses. Paper 3 especially contributes to this discussion.

8 References

Åfarli, 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

Åfarli, Tor A. 2000. *Grammatikk – kultur eller natur? Elementær innføring i det generative grammatikkstudiets vitenskapsteori. [Grammar – culture or nature? Elementary introduction in Generative Grammar's theory of science].* Oslo: Det norske samlaget.

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

Abels, Klaus & Bentzen, Kristine. 2009. A Note on the Punctuated Nature of Movement Paths. *Catalan Journal of Linguistics; Vol.: 8 Spelling-Out Universal Grammar*, 8(1), 19-40. doi: https://doi.org/10.5565/rev/catjl.138

Adger, David. 2003. *Core Syntax. A Minimalist Approach.* Oxford: Oxford University Press.

Adger, David. 2013. Constructions and Grammatical Explanation: Comments on Goldberg. *Mind & Language*, 28(4), 466-478.

Adger, David. 2021. The Architecture of the Computation. In Nicholas Allott, Terje Lohndal & Georges Rey (Eds.), *A Companion to Chomsky,* 125-139. Hoboken: Wiley Blackwell.

Allott, Nicholas. & Lohndal, Terje. (forthcoming). Minimalism: Emergence and theoretical foundations. In Kleanthes Grohmann & Evelina Leivada (Eds.), *Cambridge Handbook of Minimalism*. Cambridge: Cambridge University Press.

Ambridge, Ben & Goldberg, Adele. E. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics,* 19(3), 349–381. doi: https://doi.org/10.1515/COGL.2008.014

Anward, Jan. 1982. Basic Swedish. In Elisabet Engdahl & Eva Ejerhed (Eds.), *Readings on unbounded dependencies in Scandinavian languages*, 47-75. Stockholm: Almqvist & Wiksell.

Barlaz, Marissa. 2022. Ordinal logistic regression in R. Retrieved from: https://marissabarlaz.github.io/portfolio/ols/

Bermingrud, Knut O. 1979. Setningsknutekonstruksjonen: en analyse av konstruksjonens grammatikalitet i moderne norsk. [The «sentence-knot-construction»: an analysis of the construction's grammaticality in modern Norwegian]. MA thesis: University of Oslo.

Bhatt, Rajesh & Pancheva, Roumyana. 2006. Conditionals. In Martin Everaert & Henk van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (Vol 1)*,* 638-688. Oxford: Blackwell Publishing Ltd.

Biskup. Petr & Šimík, Radek. 2019. Structure of conditional and (cor)relative clauses: New evidence from locality. *Proceedings of NELS 49.*

Bode, Stefanie. 2020. *Casting a Minimalist Eye on Adjuncts*. New York: Routledge.

Boeckx, Cedric. 2008a. Islands. *Language and Linguistics Compass*, 2, 151-167. doi: https://doi.org/10.1111/j.1749-818X.2007.00043.x

Boeckx, Cedric. 2008b. *Understanding Minimalist Syntax. Lessons from Locality in Long-Distance Dependencies*. Oxford: Blackwell publishing.

Boeckx, Cedric. 2012. *Syntactic Islands*. New York: Cambridge.

Boeckx, Cedric & Lasnik, Howard. 2006. Intervention and Repair. *Linguistic Inquiry* 37(1), 150-155. url: https://www.jstor.org/stable/4179355

Boeckx, Cedric & Grohmann, Kleanthes K. 2007. Remark: Putting Phases in Perspective. *Syntax,* 10(2), 204–22.

Bondevik, Ingrid. 2018. Investigating the universality of adjunct islands through formal acceptability experiments: a comparative study of English and Norwegian. MA thesis, NTNU.

Bošković, Željko. 2007. On the Locality and Motivation of Move and Agree: An Even More Minimal Theory. *Linguistic Inquiry* 38(4), 589–644.

Brody, Michael. 2002. On the Status of Representations and Derivations. In Samual Epstein & Daniel T. Seely (Eds.), *Derivation and Explanation in the Minimalist Program,* 19-33. Malden Mass: Blackwell.

Brown, Jessica M. M. 2015a. Apparently semantically-motivated extraction in an autonomous syntax. *Cambridge occasional papers in linguistics,* 8, 6-22.

Brown, Jessica M. M. 2015b. Blackholes and subextraction from adjuncts in English and Norwegian. *Proceedings of CLS 51 (2015)*, 67-81.

Chatzopoulou, Katerina. 2021. The history of Greek conditionals & one modal cycle: Upward reanalysis in the life of *an*. *Journal of Historical Syntax,* 5(19), 1-47. doi: https://doi.org/10.18148/hs/2021.v5i16-25.71

Chaves, Rui & Putnam, Mike. 2020. *Unbounded dependency construction: Theoretical and experimental perspectives.* Oxford: Oxford University Press.

Chomsky, Noam. 1957/2002. *Syntactic Structures*. Mouton de Gruyer: Berlin.

Chomsky, Noam. 1959/1967. A Review of B. F. Skinner's Verbal Behavior. In Leon. A.

Jakobovits & Murray S. Miron (Eds.), *Readings in the Psychology of Language*. New Jersey: Prentice-Hall.

Chomsky, Noam. 1964. *Current Issues in Linguistic Theory*. Vol. 38. Hague: De Gruyter Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In Morris Halle & Stephen R. Anderson & Paul Kiparsky (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. 1980. *Rules and representations.* New York: Columbia University Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1986a. *Barriers* (Vol. 13). Cambridge, Mass: MIT Press.

Chomsky, Noam. 1986b. *Knowledge of language: its nature origin and use.* New York: Praeger.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Kenneth Hale & Samuel Jay Keyser (Eds.), *The View From Building 20: Essays in Linguistics in Honor of Sylvain Bromberger,* 1-52. Cambridge Mass: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, Mass: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin & David Michaels & Juan Uriagereka & Samuel J. Keyser (Eds.), *Step by step—essays on minimalist syntax in honor of Howard Lasnik*, 89–155. MIT.

Chomsky, Noam. 2001. Derivation by Phase. In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language*, 1-53. Massachusetts: MIT.

Chomsky, Noam. 2004. Beyond Explanatory Adequacy. In Adriana Belletti (Ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, Vol. 3, 104–31. New York: Oxford University Press.

Chomsky, Noam & Lasnik, Howard. 1977. Filters and control. *Linguistic Inquiry* 8(3), 425-504.

Christensen, Rune H. B. 2022. ordinal - Regression Models for Ordinal Data. R package version 2022.11-16. url: https://cran.r-project.org/package=ordinal.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. New York: Oxford University Press.

Cowart, Wayne. 1997. *Experimental Syntax: Applying Objective Methods to Sentence*

Judgments. Thousand Oaks: SAGE Publications.

Dal Farra, Chiara. 2020. To be or not to be an island – the status of adjuncts. Doctoral dissertation, Ca' Foscari Univserity of Venice.

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

Deane, Paul Douglas. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguitics,* 2(1), 1-64. doi: https://doi.org/10.1515/cogl.1991.2.1.1

den Besten, Hans. 1981. On the interaction of root transformations and lexical deletive rules. In Werner Abraham (Ed.) *On the Formal Syntax of the Westgermania*, 47-133. Amsterdam/Philiadelphia: John Benjamins Publishing Company.

den Dikken, Marcel. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Cambridge, Mass: MIT Press.

den Dikken, Marcel. 2009. On the Nature and Distribution of Successive Cyclicity. Retrieved from: https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Programs/Linguistics/Dikken/distribution_successive_cyclicity.pdf

den Dikken, Marcel & Lahne, Antje. 2013. The locality of syntactic dependencies. In Marcel den Dikken (Ed.), *The Cambridge Handbook of Generative Syntax*, 655 - 698. Cambridge: Cambridge University Press.

Dickson, Niels & Pearl, Lisa & Futrell, Richard. 2022. Learning constraints on wh-dependencies by learning how to efficiently represent wh-dependencies: A developmental modeling investigation with Fragment Grammars. *Proceedings on the Society for Computation in Linguistics 2022.* url: https://aclanthology.org/2022.scil-1.21

Drummond, Alex. 2012. IbexFarm. Retrieved from http://spellout.net/ibexfarm/

Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy,* 6, 5-34.

Epstein, Samuel D. & Seely, T. Daniel. 2002a. Introduction: On the Quest for Explanation. In Samual Epstein & Daniel T. Seely (Eds.), *Derivation and Explanation in the Minimalist Program,* 1-10. Malden Mass: Blackwell.

Epstein, Samuel D. & Seely, T. Daniel. 2002b. Rule applications as cycles in a level-free syntax. In Samual Epstein & Daniel T. Seely (Eds.), *Derivation and Explanation in the Minimalist Program,* 65-84. Malden Mass: Blackwell.

Ernst, Thomas. 2022. The adjunct condition and the nature of adjuncts. *The Linguistic Review,* 39(1), 85-128. doi: https://doi.org/10.1515/tlr-2021-2082

Erteschik-Shir, Nomi. 1973. On the nature of island constraints. Doctoral dissertation, MIT.

Faarlund, Jan T. 1992. *Norsk syntaks i et funksjonelt perspektiv. [Norwegian syntax in a functional perspective].* Oslo: Universitetsforlaget.

Featherston, Sam. Universals and grammaticality: wh-constraints in German and English. *Linguistics*, 43(4), 667-711. doi: https://doi.org/10.1515/ling.2005.43.4.667

Freidin, Robert & Lasnik, Howard. 2011. Some roots in minimalism in generative grammar. In Cedric Boeckx (Ed.) *The Oxford Handbook of Linguistic Minimalism*, 1-26. Oxford: Oxford University Press.

Gallego, Angel. 2012. Introduction: A Framework of Phases for Linguistic Theory. In Angel Gallego (Ed.), *Phases - developing a framework*, 9–43. Boston: De Gruyter Mouton.

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

Goldberg, Adele E. 2013. Backgrounded constitutents cannot be "extracted". In Jon Sprouse & Norbert Hornstein (Eds.), *Experimental Syntax and Island Effects*, 221-238. Cambridge: Cambridge University Press.

Goodall, Grant. 2021. Sentence Acceptability Experiments: What, How, and Why. In Grant Goodall (Ed.), *The Cambridge Handbook of Experimental Syntax*, 7-38.

Gross, Steven. 2021. Linguistic Judgments as Evidence. In Nicholas Allott & Terje Lohndal & Georges Rey (Eds.), *A Companion to Chomsky*, 544-556. Hoboken: John Wiley & Sons, Inc.

Haarstad, Stian & Lohndal, Terje & Mæhlum, Brit. 2017. *Innganger til språkvitenskap: teori, metode og faghistorie. [Entryways to linguistic science: theory, method and history of the field]*. Oslo: Cappelen Damm akademisk.

Haegeman, Liliane. 2010. The Movement Derivation of Conditional Clauses. *Linguistic Inquiry* 41(4), 595-621. url: https://www.jstor.org/stable/40926397

Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena, and composition of the left periphery*. New York: Oxford University Press.

Henry, Alison. 1995. Inversion in Embedded Questions. In *Belfast English and Standard English: Dialect Variation and Parameter Setting.* New York: Oxford University Press.

Henry, Alison. 2005. Non-standard dialects and linguistic data. *Lingua, 115*(11), 1599-1617. doi: https://doi.org/10.1016/j.lingua.2004.07.006

Hofmeister, Philip & Sag, Ivan A. 2010. Cognitive Constraints and Island Effects. *Language,* 86(2), 366-415. url: http://www.jstor.org/stable/40666324

Hofmeister, Philip & Casanto, Laura S. & Sag, Ivan A. 2013. Islands in the grammar? Standards of evidence. In Jon Sprouse & Norbert Hornstein (Eds.), *Experimental Syntax and Island Effects*, 64–108. Cambridge: Cambridge University Press.

Holmberg, Anders. 2015. Verb Second. In Tibor Kiss & Artemis Alexiadou (Eds.), *Syntax – Theory and Analysis: An International Handbook,* 241-270. Berlin: De Gruyter Mouton.

Hornstein, Norbert & Lightfoot, David. 1981. *Explanation in linguistics. The logical problem of language acquisition*. New York: Longman Group Limited.

Hornstein, Norbert & Nunes, Jairo & Grohmann, Kleanthes K. 2005. *Understanding Minimalism.* Cambridge: Cambridge University 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. Doctoral dissertation, MIT.

Joseph, John E. 2010. Chomsky's Atavistic Revolution (with a little help from his enemies). In Douglas A. Kibbee (Ed.), *Chomskyan (R)evolutions*. Amsterdam: John Benjamins Publishing Company.

Kim, Boyoung & Goodall, Grant. 2016. Islands and Non-islands in Native and Heritage Korean. *Frontiers in Psychology*, 7, 134. doi: https://doi.org/10.3389/fpsyg.2016.00134

Kim, Ilkyu. 2021. A Note on the Factorial Definition of Island Effects. *Language Research* 57(2), 211-223. doi: https://doi.org/10.30961/lr.2021.57.2.211

Kluender, Robert & Kutas, Marta. 1993. Bridging the Gap: Evidence from ERPs on the Processing of Unbounded Dependencies. *Journal of Cognitive Neuroscience,* 5(2), 196-214.

Klævik-Pettersen, Espen. 2022. Full V2, no V2, residual V2: Exploring variation through phases. *Isogloss* 8(3)/5, 1-25. doi: https://doi.org/10.5565/rev/isogloss.199

Kobzeva, Anastasia & Arehalli, Suhas & Linzen, Tal & Kush, Dave. 2023. Neural Networks Can Learn Patterns of Island-insensitivity in Norwegian. *Proceedings of the Society for Computation in Linguistics (SCiL)* 2023, 175-185. url: https://aclanthology.org/2023.scil-1.15.pdf

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. *Procedings of the Linguistic Society of America*, 5(1), 807–821. doi: https://doi.org/10.3765/plsa.v5i1.4775

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: 10.1353/lan.0.0237

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 & Sant, Charlotte & Strætkvern, Sunniva. 2021. Learning Island-insensitivity from the input: A corpus analysis of child- and youth-directed text in Norwegian. *Glossa: a journal of general linguistics*, 6(1), 105. doi: https://doi.org/10.16995/glossa.5774

Lasnik, Howard. 2001. Derivation and Representation in Modern Transformational Syntax. In Mark Baltin and Chris Collins (Eds.), *The Handbook of Contemporary Syntactic Theory*. doi: https://doi.org/10.1002/9780470756416.ch3

Lasnik, Howard. 2003. On the Extended Projection Principle. Retrieved from: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.6842&rep=rep1&type=pdf

Lasnik, Howard. 2006. Conceptions of the Cycle. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Wh movement: Moving On*, 197–216. Cambridge, Mass: MIT Press.

Lasnik, Howard & Uriagereka, Juan. 1988. *A course in GB syntax: lectures on binding and empty categories.* Cambridge, Mass: MIT Press.

Lasnik, Howard & Lohndal, Terje. 2013. Brief overview of the history of generative syntax. In Marcel den Dikken (Ed.), *The Cambridge Handbook of Generative Syntax,* 26-60. Cambridge: Cambridge University Press.

Lasnik, Howard & Lidz, Jeffrey L. 2017. The Argument from the Poverty of the Stimulus. In Ian Robert (Ed.), *The Oxford Handbook of Universal Grammar*. Oxford: Oxford University Press.

Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical*

analysis. Amsterdam: John Benjamins Publishing Company.

Liejiong, Xi. 1990. Are they parasitic gaps? In Joan Mascaró & Marina Nespor (Eds.), *Grammar in Progress: GLOW Essays for Hank Van Riemsdijk*, 455-461. Dordrecht: De Gruyter Mouton.

Lohndal, Terje. 2009. *Comp*-t Effects. Variation in the Position and Features of C. *Studia Linguistica* 63(2), 204-232. doi: https://doi.org/10.1111/j.1467-9582.2009.01159.x

Lohndal, Terje. 2018. Introduction. In Terje Lohndal (Ed.) *Formal Grammar. Theory and Variation across English and Norwegian.* Routledge: New York.

Lüdecke, Daniel. 2018. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software,* 3(26), 772. doi: https://doi.org/10.21105/joss.00772

MacCorquodale, Kenneth. 1970. On Chomsky's Review of Skinner's Verbal Behavior. *Journal of the Experimental Analysis of Behavior*, 13(1), 83-99.

Marty, Paul & Chemla, Emmanuel & Sprouse, Jon. 2020. The effect of three basic task features on the sensitvity of accetpability judgment tasks. *Glossa: a journal of general linguistics,* 5(1), 72. doi: https://doi.org/10.5334/gjgl.980

McCloskey, James. 2000. Quantifier Float and *Wh*-Movement in an Irish English. *Linguistic Inquiry*, 31(1), 57-84.

McInnery, Andrew R. 2022. The Argument/Adjunct Distinction and the Structure of Prepositional Phrases. Doctoral dissertation, the University of Michigan.

Munaro, Nicola (2005). Computational puzzles of conditional clause preposing. In Anna Maria Di Sciullo (Ed.), *UG and External Systems: Language, Brain and Computation,* 71-94. Philadelphia: John Benjamins Publishing Co.

Müller, Christiane. 2019. Permeable islands. Doctoral dissertation, Lund University.

Müller, Christiane & Eggers, Clara U. 2022. Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages* 7(2), 125. doi: https://doi.org/10.3390/languages7020125

Myers, James. 2012. Testing Adjunct and Conjunct Island Constraints in Chinese. *Language and Linguistics,* 13(3), 437-470.

Namboodiripad, Savithry & Bisnath, Felicia & Kramer, Alex & Luntzlara, Noah & Goldberg Adele E. 2021. Backgroundedness measures predict island status of non-finite adjuncts in English. Poster, CUNY 2021.

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

Pearl, Lisa & Sprouse, Jon. 2013. Computational models of acquisition for islands. In Jon Sprouse & Norbert Hornstein (Eds.), *Experimental Syntax and Island Effects*, 109–131. Cambridge: Cambridge University Press.

Perlmutter, David M. 1968. Deep and surface structure constraints in syntax. Doctoral dissertation, MIT.

Phillips, Colin. 2013a. On the nature of island constraints I: Language processing and reductionist accounts. In Jon Sprouse & Norbert Hornstein (Eds.), *Experimental Syntax and Island Effects*, 64–108. Cambridge: Cambridge University Press.

Phillips, Colin. 2013b. On the nature of island constraints II: Language learning and innateness. In Jon Sprouse & Norbert Hornstein (Eds.), *Experimental Syntax and Island Effects,* 132–158. Cambridge: Cambridge University Press.

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/.

Richards, Norvin. 2011. A-Bar Dependencies. In Cedric Boeckx (Ed.), *The Oxford Handbook of Linguistic Minimalism,* 173–94. New York: Oxford University Press.

Rizzi, Luigi. 1990. *Relativized Minimality*. Linguistic Inquiry Monographs (16). Cambridge Mass: MIT.

Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In Liliane Haegeman (Ed.) *Elements of Grammar. Handbook in Generative Syntax*, 281-338. Dordrecht: Springer.

Rizzi, Luigi. 2001. Relativized Minimality Effects. In Mark Baltin & Chris Collins (Eds.) *The Handbook of Contemporary Syntactic Theory*, 89-110. doi: https://doi.org/10.1002/9780470756416.ch4

Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In Lisa Lai-Shen Cheng & Norbert Corver (Eds.), *Wh-movement: Moving on*, 97–134. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2013. Locality. *Lingua,* 130, 169-186. doi: http://dx.doi.org/10.1016/j.lingua.2012.12.002

Rizzi, Luigi. 2018. Intervention effects in grammar and language acquisition. *Probus,* 30(2), 339-367. doi: https://doi.org/10.1515/probus-2018-0006

Rizzi, Luigi & Shlonsky, Ur. Strategies of subject extraction. In Uli Sauerland & Hans-Martin

Gärtner (Eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*, 115-160. Berlin: Mouton de Gruyter.

Rizzi, Luigi & Bocci, Giuliano. 2017. Left Periphery of the Clause. Primarily Illustrated for Italian. In Martin Everaert & Henk Van Riemsdijk (Eds.), *The Companion to Syntax*. John Wiley and Sons, Inc. doi: https://doi.org/10.1002/9781118358733.wbsyncom104

Ross, John R. 1967. Constraints on variables in syntax. Doctoral dissertation, MIT.

Rubin, Edward J. 2003. Determining Pair-Merge. *Linguistic Inquiry (Scuibs and Discussion),* 34(4), 660-668.

Schütze, Carson T. 2020. Acceptability ratings cannot be taken at face value. In Samuel Schindler & Anna Drożdżowicz & Karen Brøcker (Eds.) *Linguistic Intuitions – Evidence and Method,* 189-214. Oxford: Oxford University Press.

Schütze, Carson T. & Sprouse, Jon. 2014. Judgment Data. In Robert J. Podesva and Devyani Sharma (Eds.), *Research Methods in Linguistics.* Cambridge: Cambridge University Press.

Sprouse, Jon. 2007. A program for experimental syntax. Doctoral dissertation, University of Maryland.

Sprouse, Jon & Fukuda, Shin & Ono, Hajime & Kluender, Robert. 2011. Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax*, 14(2), 179–203. doi: https://doi.org/10.1111/j.1467-9612.2011.00153.x

Sprouse, Jon & Wagers, Matt & Phillips, Colin. 2012a. A Test of the Relation between Working-Memory Capacity and Syntactic Island Effects. *Language*, 88(1), 82–123. doi: 10.1353/lan.2012.0004

Sprouse, Jon & Wagers, Matt & Phillips, Colin. 2012b. Working-memory capacity and island effects: A reminder of the issues and the facts. *Language,* 88(2), 401-407. url: https://www.jstor.org/stable/23251839

Sprouse, Jon & Hornstein, Norbert. 2013. Experimental syntax and island effects: Toward a comprehensive theory of islands. In Jon Sprouse & Norbert Hornstein (Eds.), *Experimental Syntax and Island Effects*, 1–18. Cambridge: Cambridge University Press.

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

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 & Almeida, Diogo. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: a journal of general linguistics* 2(1), 14. doi: https://doi.org/10.5334/gjgl.236

Sprouse, Jon & Villata, Sandra. 2021. Island effects. In Goodall, Grant (Ed.), *The Cambridge Handbook of Experimental Syntax*, 227–257. Padstow: Cambridge University Press.

Starke, Michal. 2001. Move Dissolves into Merge: a Theory of Locality. Doctoral dissertation, University of Geneva.

Stepanov, Arthur. 2001. Late Adjunction and Minimalist Phrase Structure. *Syntax,* 2(4), 94-125. doi: https://doi.org/10.1111/1467-9612.00038

Stepanov, Arthur. 2007. The End of CED? Minimalism and Extraction Domains. *Syntax*, 10(1), 80–126. doi: 10.1111/j.1467-9612.2007.00094.x

Svenonius, Peter. 2021. Merge and Features: The Engine of Syntax. In Nicholas Allott, Terje Lohndal & Georges Rey (Eds.), *A Companion to Chomsky,* 140-157. Hoboken: Wiley Blackwell.

Szabolcsi, A., & Lohndal, T. 2017. Strong vs. Weak Islands. In Martin Everaert & Henk Van Riemsdijk (Eds.), *The Companion to Syntax*, 1–51. John Wiley and Sons, Inc. doi: https://doi.org/10.1002/9781118358733.wbsyncom008

Tanaka, Misako. 2020. Similarities and Differences between Quantifier Raising and Wh Movement Out of Adjuncts. *Syntax,* 23(2), 152-184. doi: https://doi.org/10.1111/synt.12189

Torr, John. 2012. Phasing out Phases and Re-Relativizing Relativized Minimality. *Cambridge occasional papers in linguistics,* 6(4), 95-129.

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.

Villata, Sandra & Rizzi, Luigi & Franck, Julie. 2016. Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua*, 179, 76–96. doi: http://dx.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.

Wilcox, Ethan Gotlieb, Richard Futrell, and Roger Levy. 2023. Using Computational Models to Test Syntactic Learnability. *Linguistic Inquiry*, 1-44. doi: https://doi.org/10.1162/ling_a_00491

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 Maria LaMendola & Jennifer Scott (Eds.), *Proceedings of the 41st annual Boston University Conference on language development*, 806–819. Somerville: Cascadilla Press.

Paper 1

Nordic Journal of Linguistics (2020), page 1 of 32
doi:10.1017/S0332586520000207

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Variation in adjunct islands: The case of Norwegian

Ingrid Bondevik[1], Dave Kush[1] and Terje Lohndal[1,2]

[1]Department of Language and Literature, NTNU, 7491 Trondheim, Norway and [2]UiT The Arctic University of Norway
Email for correspondence: ingrid.bondevik@ntnu.no, dave.kush@ntnu.no, 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

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).[1]

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,[2] 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 *sportspegeln* '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.[3] 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.[4] 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	*Short-distance*, *no-Island* (6a)	*Short-distance*, *Island* (6c)
	Long-distance	*Long-distance*, *no-Island* (6b)	*Long-distance*, *Island* (6d)

(6) *'Whether'-island example from Sprouse* (2007:56)
   a. Who _ thinks [that you wrote the letter]?
   b. What _ do you think [that you wrote _]?
   c. Who _ wonders [whether you wrote the letter]?
   d. 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)[5] score (Maxwell & Delaney 2003). This allows (mean) effect sizes to be compared across islands and experiments.

2.1.2 Materials[6]

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

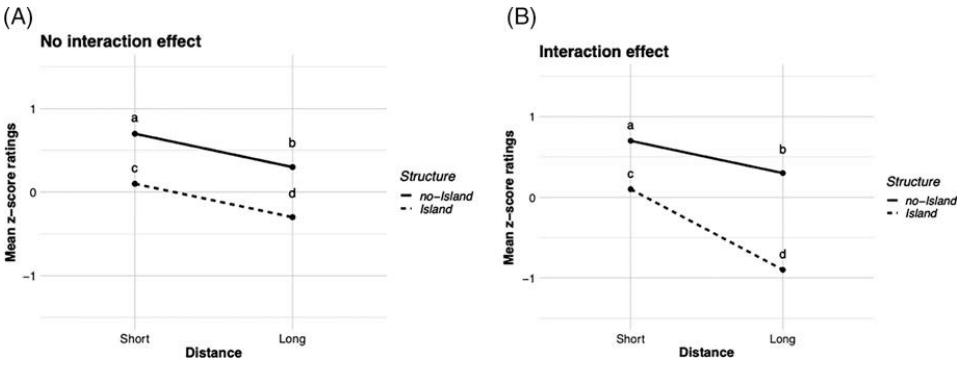


Figure 1. Example interaction plots illustrating the absence of a *Distance* × *Structure* island effect (A) or the presence of a *Distance* × *Structure* island effect (B).

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 disapppointed 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
but he thinks the new.DEF treatment.DEF against cancer deserves
Nobelprisen.
Nobel.Prize.DEF
'but he thinks the new treatment against cancer deserves the Nobel Prize.'
d. men kreft synes han den nye behandlingen mot
but cancer thinks he the new.DEF treatment.DEF against
fortjener Nobelprisen.
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).[7] 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* × *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* × *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* condtion). 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* × *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.

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

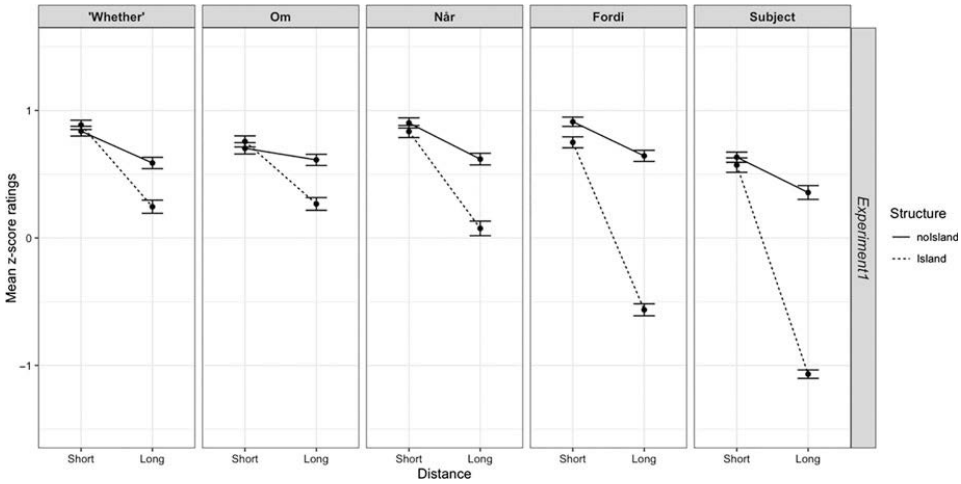


Figure 2. Interaction plots for Experiment 1. Error bars indicate standard error.

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

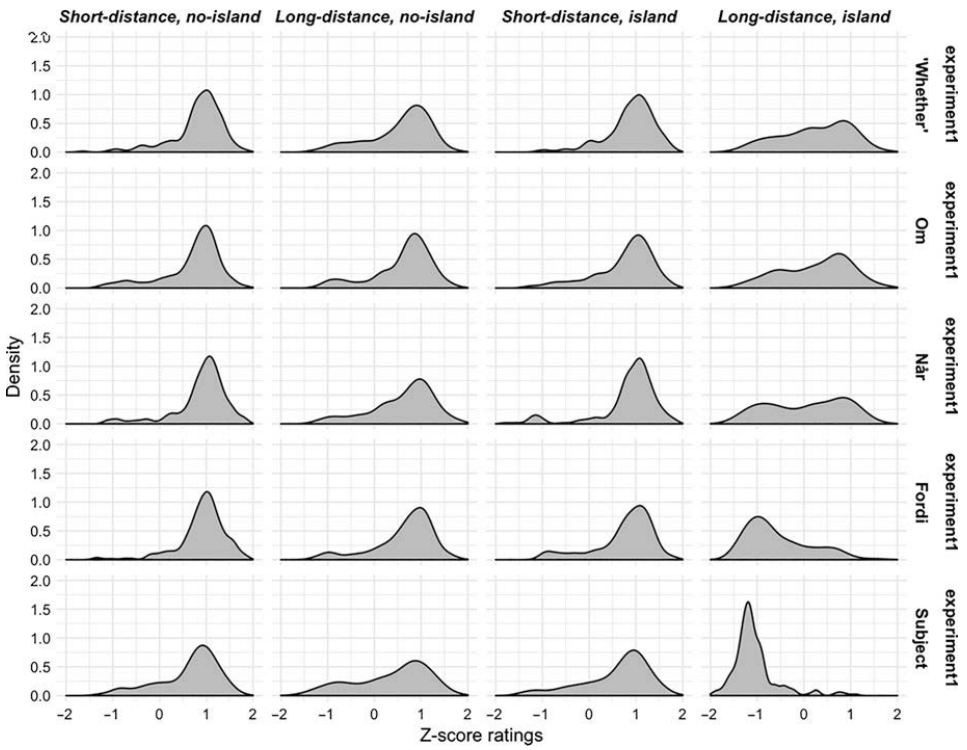


Figure 3. Distribution of *z*-scores for each island type tested and for each condition.

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.[8] 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

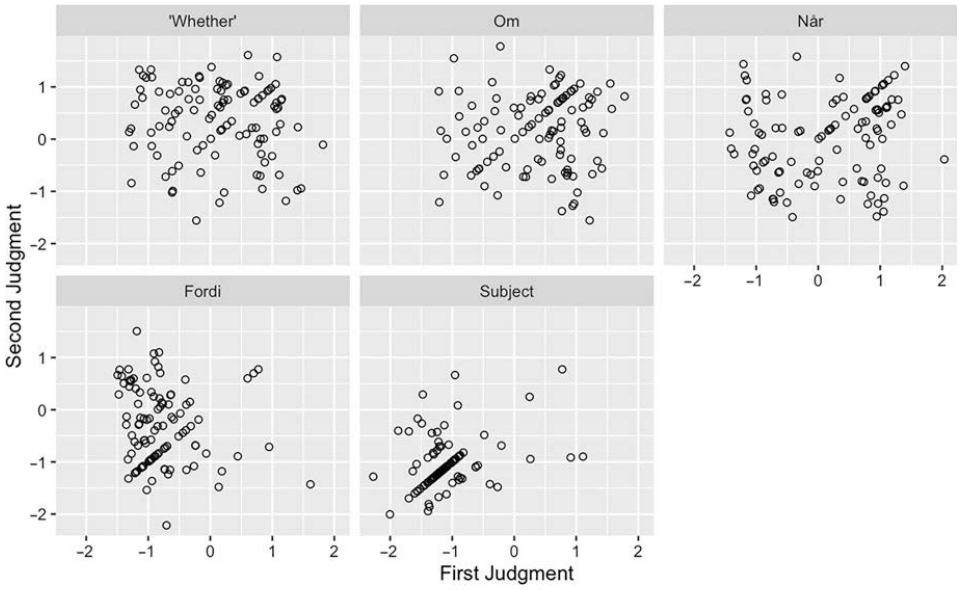


Figure 4. Each participant's judgments split by island type in Experiment 1. Each dot represents one participant, with their first judgment (*x*-axis) plotted against their second judgment (*y*-axis) on the *Long-distance*, *Island* condition.

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 lefthand 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 lefthand quadrant for subject islands in Figure 4. Many participants consistently accepted 'whether'-island

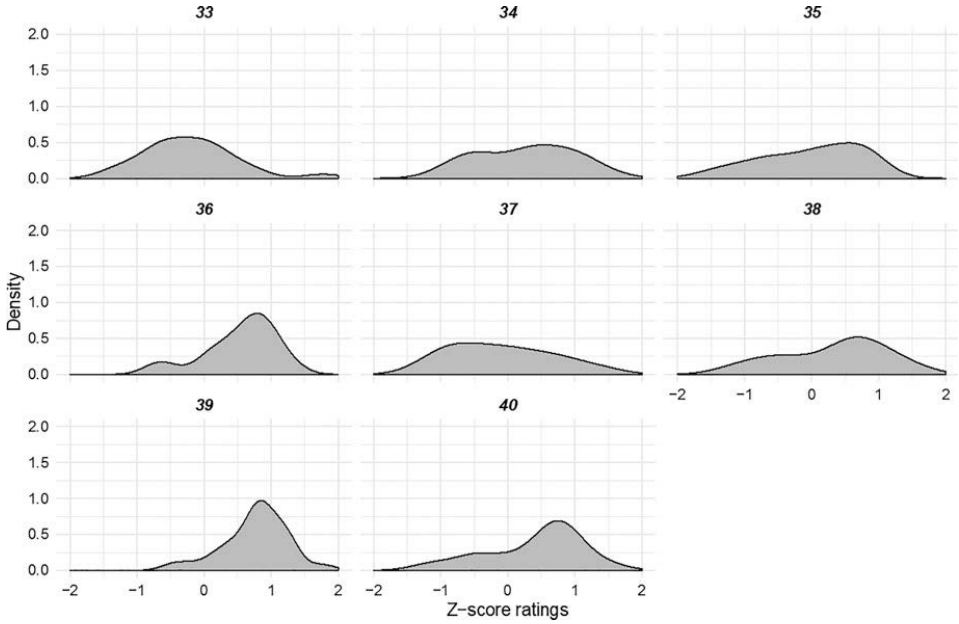


Figure 5. Distribution of *z*-scores for the *Long-distance*, *Island* condition for *om*-items tested. Item numbers are provided for cross-reference in the materials list.

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

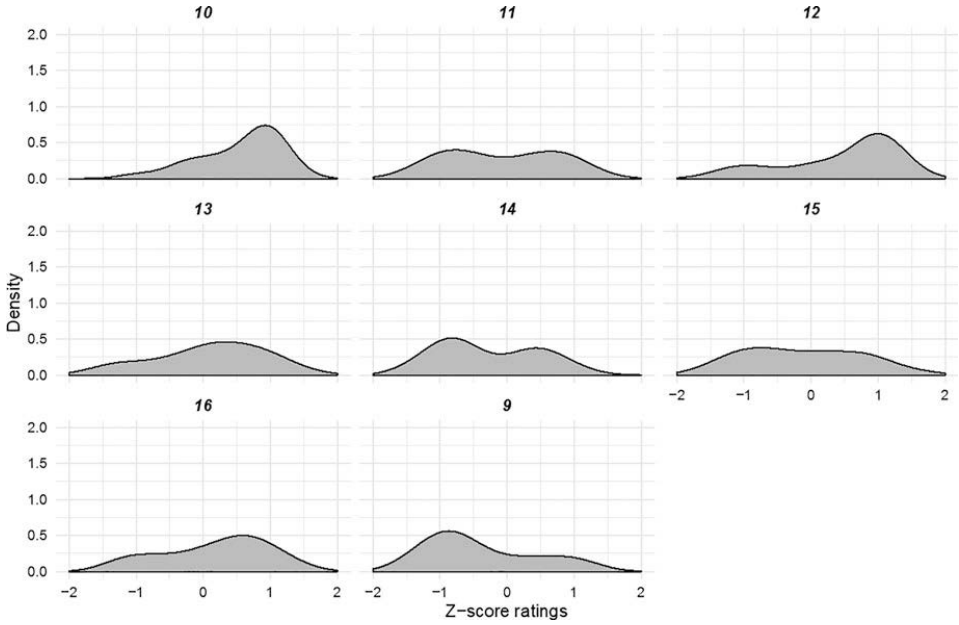


Figure 6. Distribution of z-scores for the *Long-distance, Island* condition for *når*-items tested. Item numbers are provided for cross-reference in the materials list.

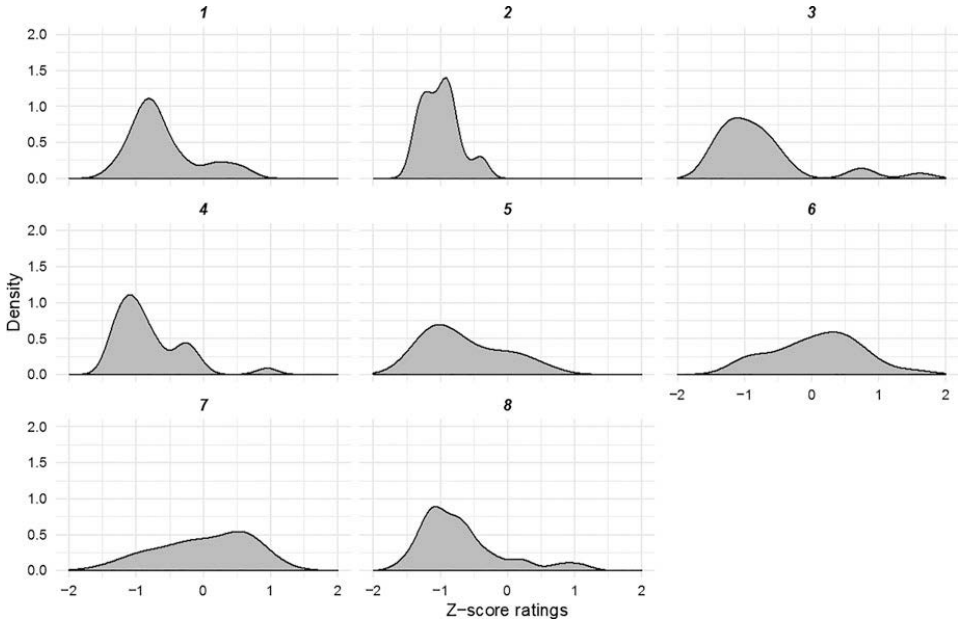


Figure 7. Distribution of *z*-scores for the *Long-distance, Island* condition for *fordi*-items tested. Item numbers are provided for cross-reference in the materials list.

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*-, *når*-, 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 *når*-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 *når*-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.[9] 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 *når*, 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 *når*.

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.[10] In addition we added four acceptable fillers featuring local topicalization to have a rough baseline of acceptability for topicalization across a single clause.

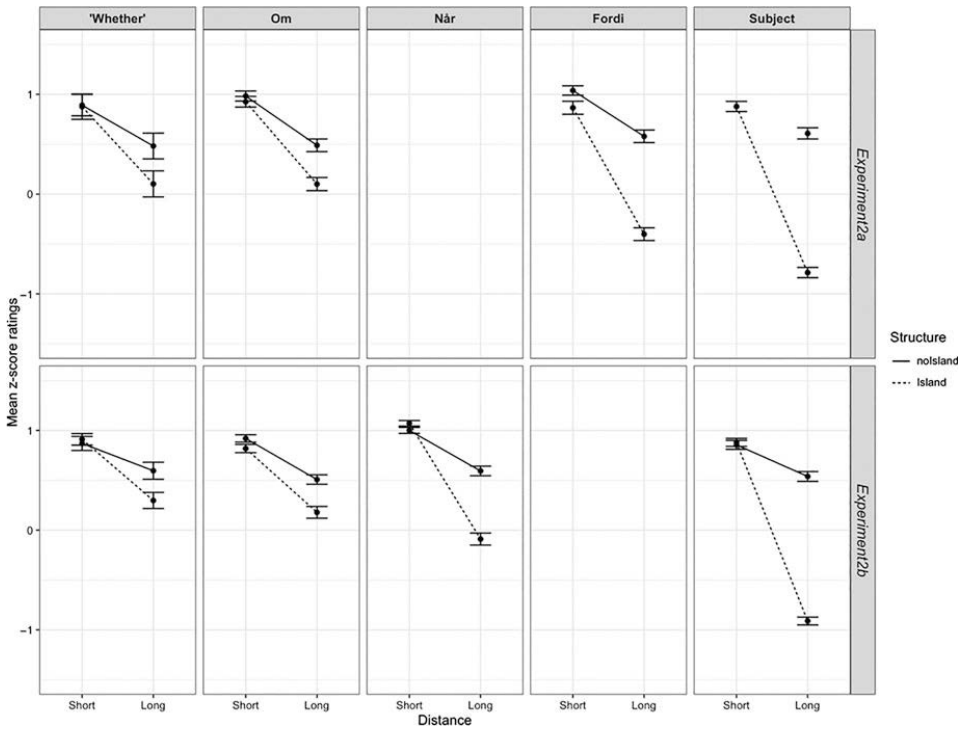


Figure 8. Interaction plots for Experiment 2a and 2b. Error bars indicate standard error.

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.

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

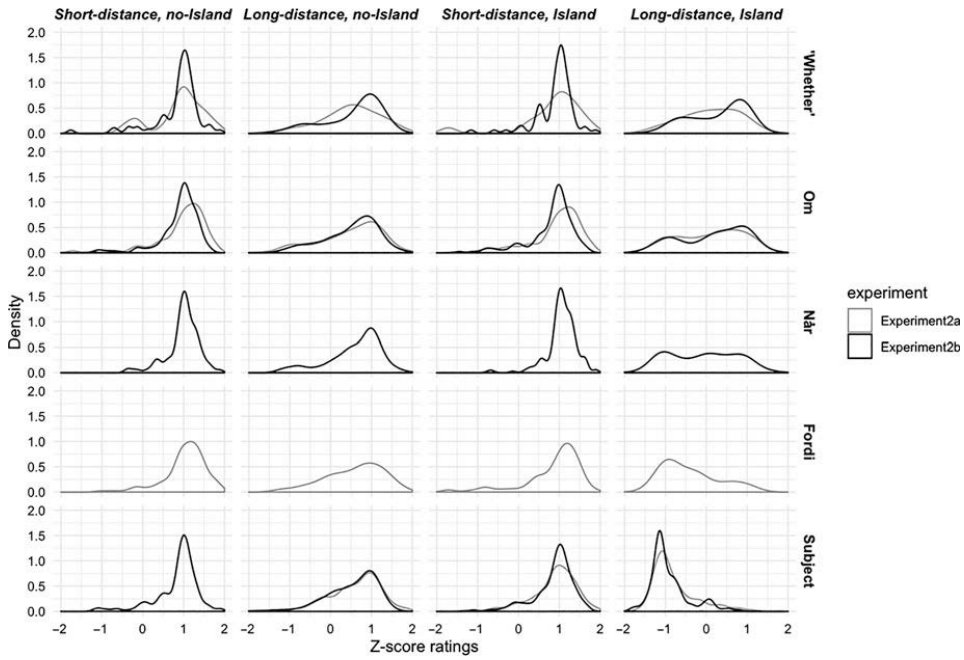


Figure 9. Distribution of *z*-scores for each condition in adjunct island comparisons in Experiments 2a and 2b.

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

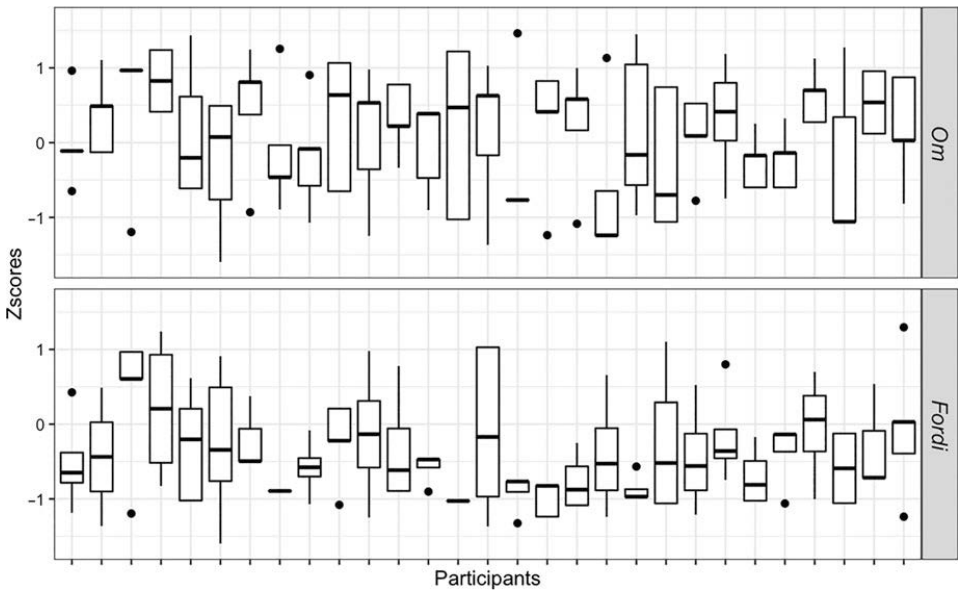


Figure 10. Overview of participant ratings of *om-* and *fordi*-adjunct items in Experiment 2a on the *Long-distance, no-Island* condition.

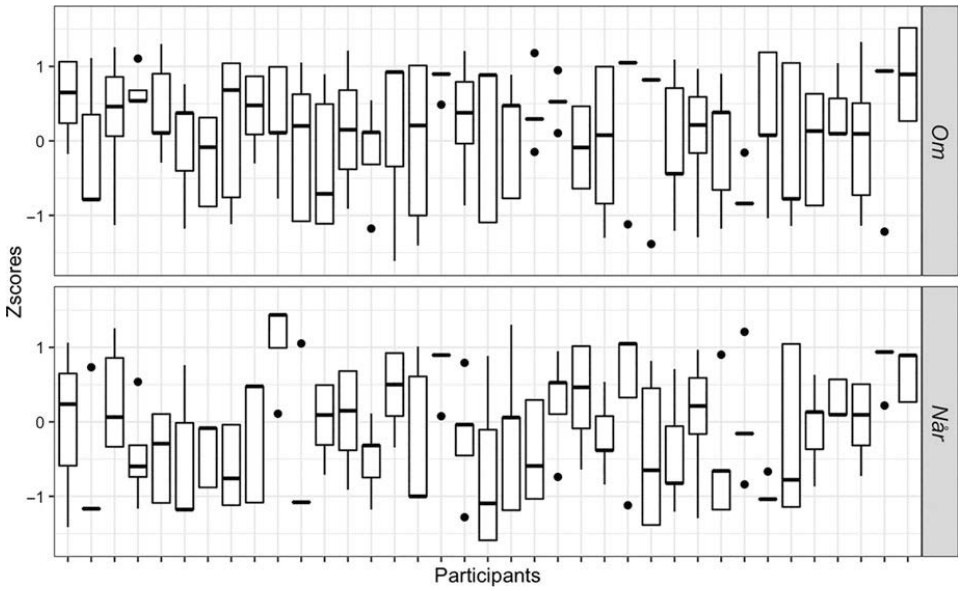


Figure 11. Overview of participant ratings of *om-* and *når*-adjunct items in Experiment 2b on the *Long-distance, no-Island* condition.

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.

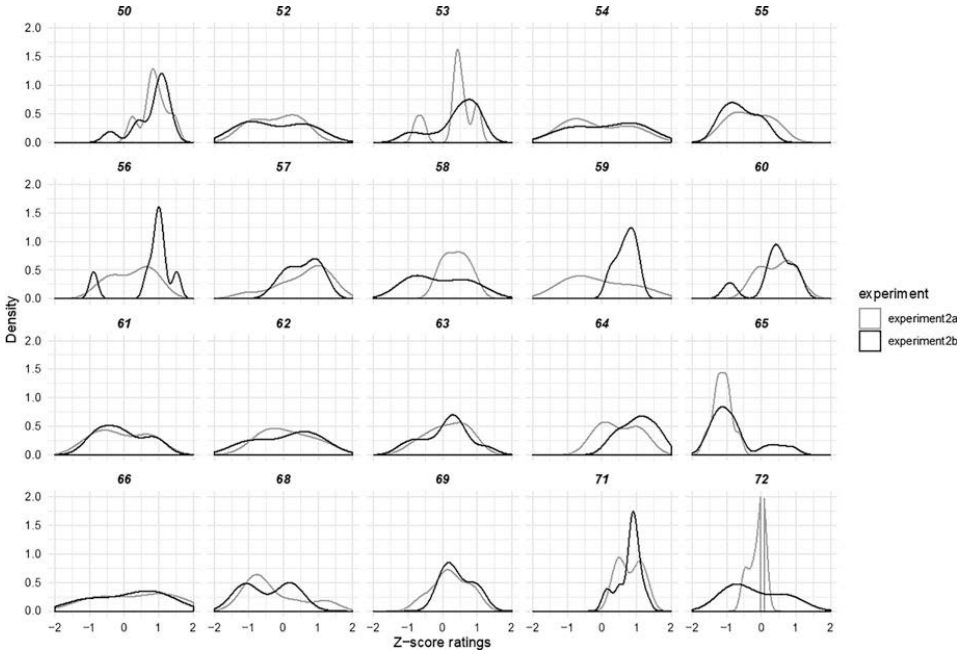


Figure 12. Distribution of *z*-scores in the *Long-distance, Island* condition for *om*-items tested in Experiments 2a and 2b. Item numbers are provided for cross-reference in the materials list.

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.

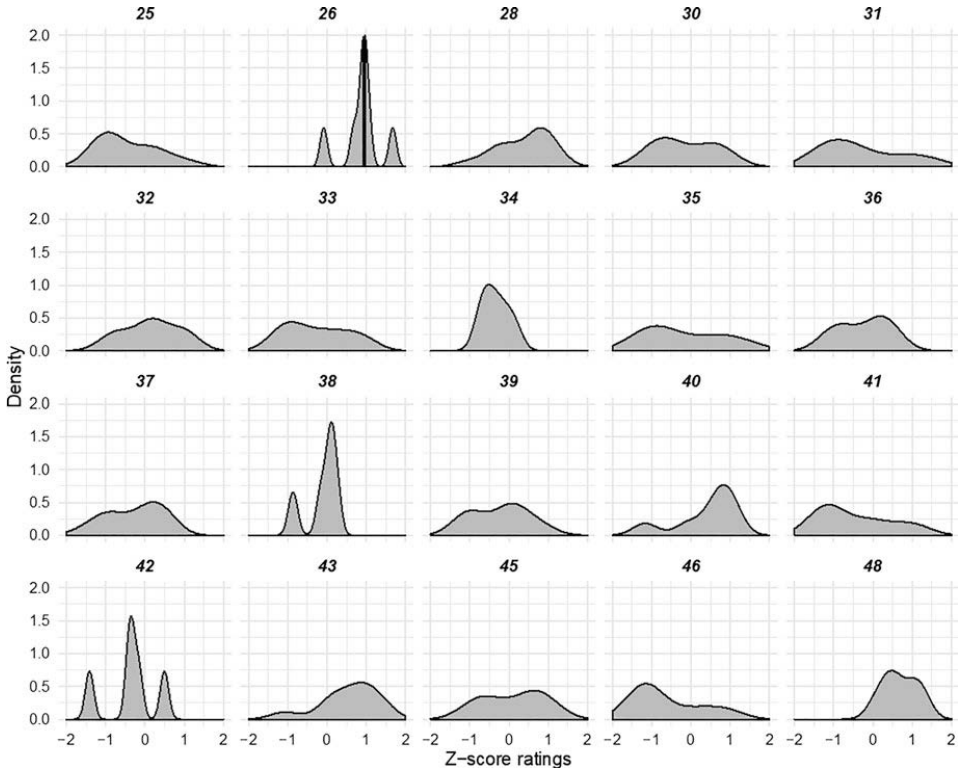


Figure 13. Distribution of *z*-scores in the *Long-distance, Island* condition for *når*-items tested in Experiment 2b. Item numbers are provided for cross-reference in the materials list.

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

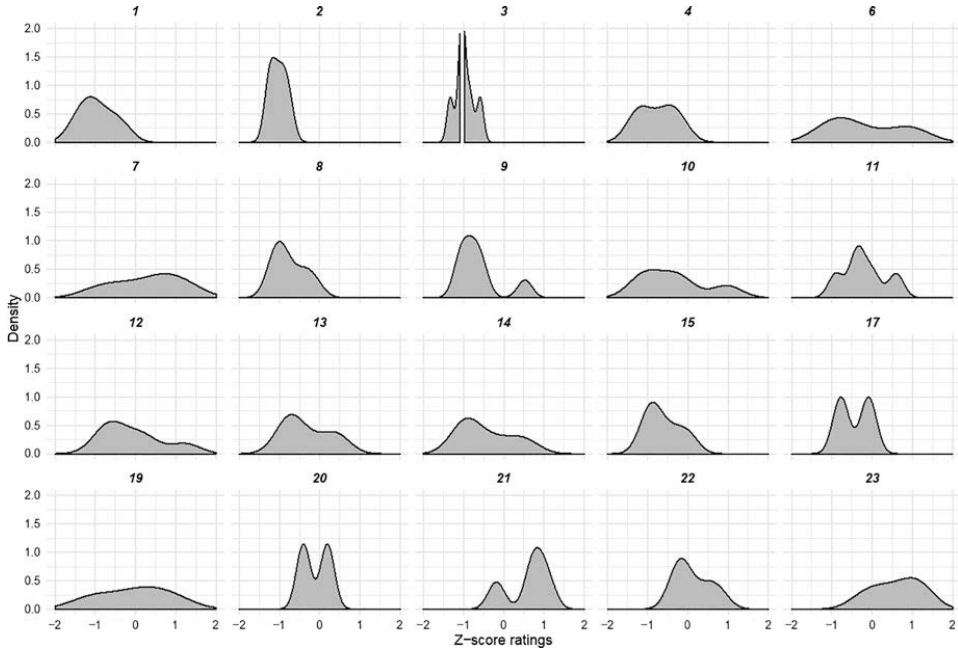


Figure 14. Distribution of *z*-scores in the *Long-distance, Island* condition for *fordi*-items tested in Experiment 2a. Item numbers are provided for cross-reference in the materials list.

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 *v*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 disapppointed 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).[11]

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.[12]

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.[13] 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 spatiotemorally – 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.[14]

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.

Aknowledgements. 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 = *Long-distance*, *no-Island* – *Long-distance*, *Island*]; (ii) calculating the difference between the short conditions [D2 = *Short-distance*, *no-Island* – *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 constrast, 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 siting 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 Univserity 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. IbexFarm. Retrieved from http://spellout.net/ibexfarm/ (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.

Cite this article: Bondevik I, Kush D, and Lohndal T. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics*. https://doi.org/10.1017/S0332586520000207

Paper 2

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

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.

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 illict (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 intermediate[6] 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 rc*s 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 rc*s and *wh*-dependencies, Kobzeva et al. (2022) find that *dem rc*s 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 rc*s 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 rc*s 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* × *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?	*short* | *no-island*
b.	What did Rachel believe that Monica bought _ ?	*long* | *no-island*
c.	Who _ was sad because Monica bought a house?	*short* | *island*
d.	What was Rachel sad because Monica had bought _?	*long* | *island*

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.

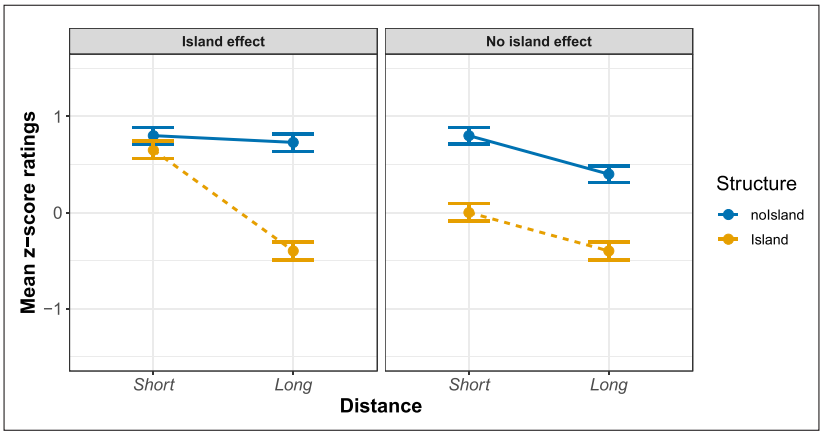


Figure 1: Examples of interaction patterns.

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 Vmatrix, Vrel and Vadjunct, 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* 'Bergen-dialect' 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 × 3 clause types = 48 adjunct items), while 8 items were tested for each of the control clause types (8 items × 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 alphabethically 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**	***t***	***p***
distance: short	–0.353	0.019	–18.369	<0.0001
structure: no-island	–0.253	0.017	–13.690	<0.0001
Interaction				
Fordi 'because'*	–0.254	0.017	–14.718	<0.0001
Når 'when'	–0.033	0.024	1.385	0.166
Om 'if'	0.140	0.024	5.747	<0.0001
Subject	–0.064	0.029	–2.148	0.0317
Whether	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 *fordi* '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 *fordi* 'because'-clauses, and similarly between the 'whether'-clauses and the *fordi* 'because'-clauses. This indicates that the interaction of distance and structure is statistically significantly different between *fordi* '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: *Fordi* '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	*t*	*p*	*DD*	*Avg. z-score: isl.cond.*
Subject				1.226	–0.605
intercept	0.377	7.581	< 0.0001		
distance	–0.328	–14.283	< 0.0001		
structure	–0.348	–15.137	< 0.0001		
distance × *structure*	–0.308	–13.406	< 0.0001		
'whether'				–0.086	0.458
intercept	0.600	13.520	< 0.0001		
distance	–0.117	–5.105	< 0.0001		
structure	–0.040	–1.760	0.079		
distance × *structure*	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	*t*	*p*	*DD*	*Avg. z-score: isl.cond.*
***Fordi* 'because'**				1.006	–0. 568
intercept	0.278	6.078	< 0.0001		
distance	–0.352	–19.084	< 0.0001		
structure	–0.254	–13.337	< 0.0001		
distance × *structure*	–0.252	–13.619	< 0.0001		
***Når* 'when'**				0.876	–0.342
intercept	0.386	6.709	< 0.0001		
distance	–0.294	–16.548	< 0.0001		
structure	–0.218	–12.223	< 0.0001		
distance × *structure*	–0.221	–12.422	< 0.0001		
***Om* 'if'**				0.469	0.082
intercept	0.489	10.211	< 0.0001		
distance	–0.109	–6.384	< 0.0001		
structure	–0.183	–10.627	< 0.0001		
distance × *structure*	–0.118	–6.870	< 0.0001		

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

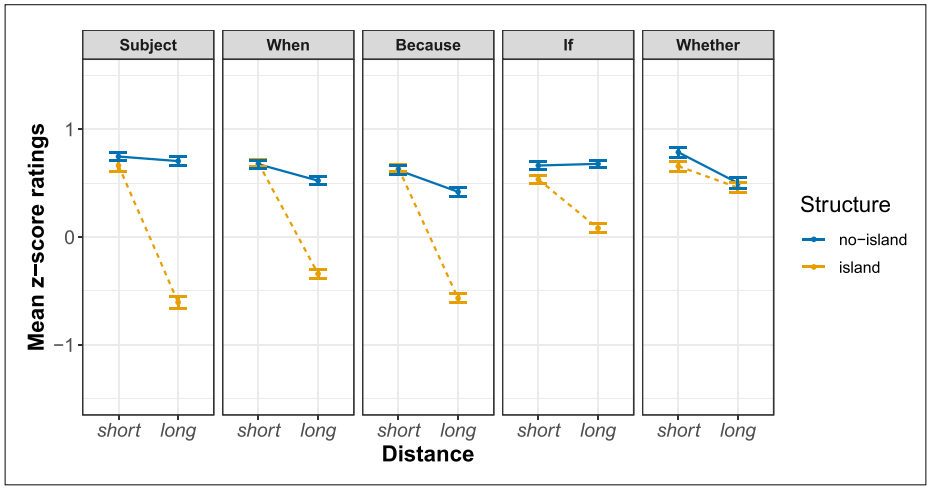


Figure 2: Interaction plot for all island types, Experiment 1 – average ratings on every condition for each clause.

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 the dgof-package (Arnold & Emerson 2011).

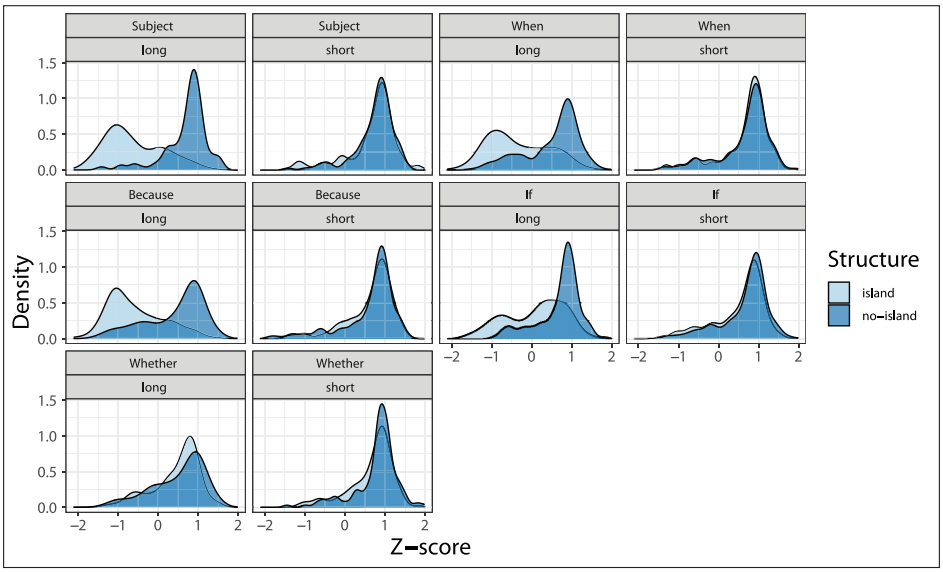


Figure 3: Comparing the distribution of z-scores on the no-island and the island conditions for the long and short conditions separately, Experiment 1.

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' ($p < 0.0001$).

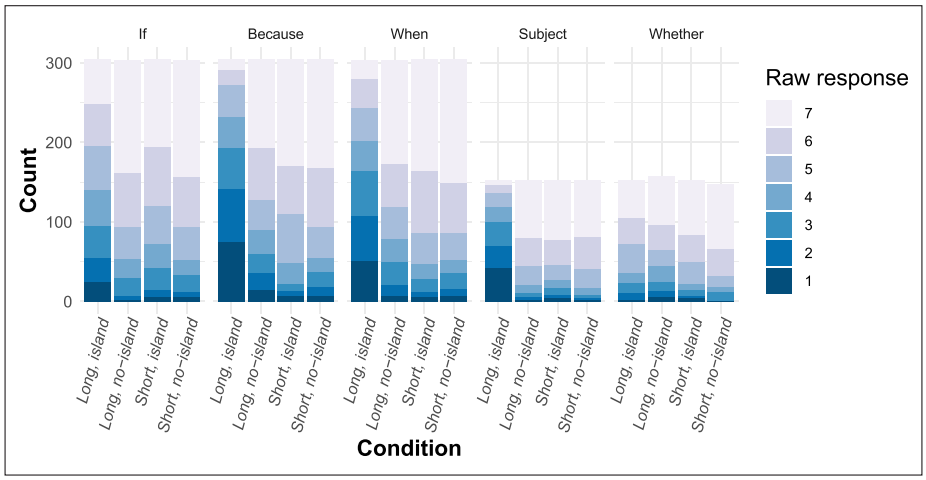


Figure 4: Barplots displaying the count of raw responses per condition for Experiment 2.

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	*t*	*p*
Fordi 'because'	–0.878	0.005	0.0001	4.928	<0.0001
Når 'when'	–0.542	0.003	0.0010	3.021	0.0027
Om '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 presentented 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.

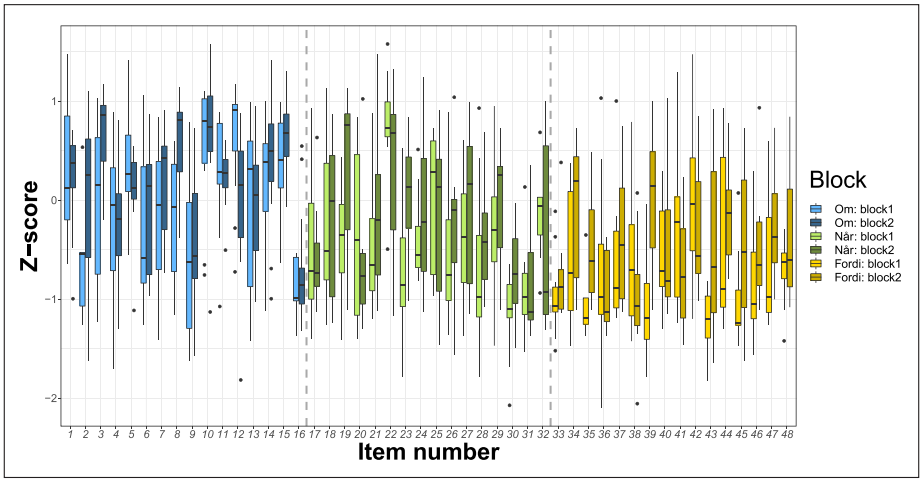


Figure 5: Boxplot illustrating average judgments on the "long, island" condition by item for each island type. The dashed line highlights the border between adjunct clause types. The plot legend provides the explanation of the colors.

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.

Condition	***Mean z-score***	***SD***
Bad fillers	–0.896	0.726
Good fillers	0.761	0.666
Short, no-island	0.736	0.569
Long, no-island	0.675	0.581
Short, island	0.484	0.673
Long, island	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*
Intercept	0.030	0.095	0.308	0.791
distance: short	0.454	0.105	4.320	<0.001
structure: no-island	0.645	0.085	7.558	<0.0001
distance × *structure*	–0.392	0.106	–3.680	0.002

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

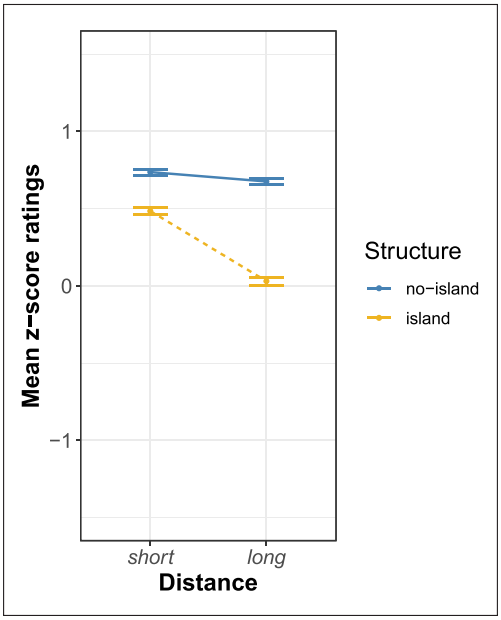


Figure 6: Interaction plot for *om* 'if', Experiment 2.

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 & Calcagnì 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.

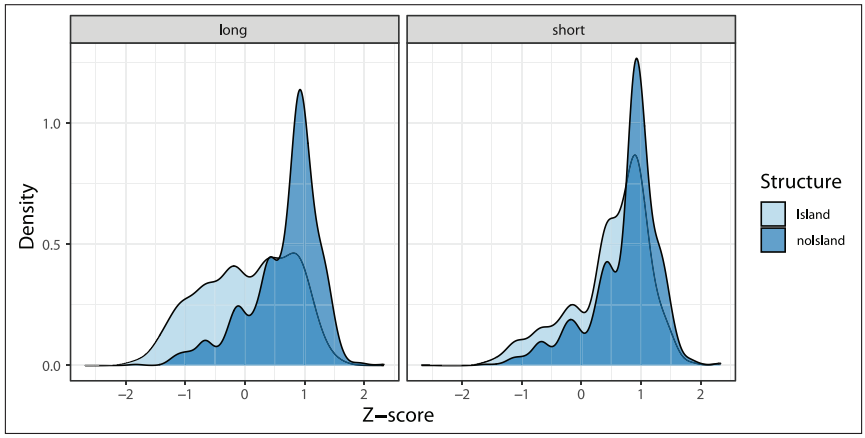


Figure 7: Comparing the distribution of z-scores for *om* 'if' on the no-island and the island conditions for the long and short conditions separately, Experiment 2.

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.

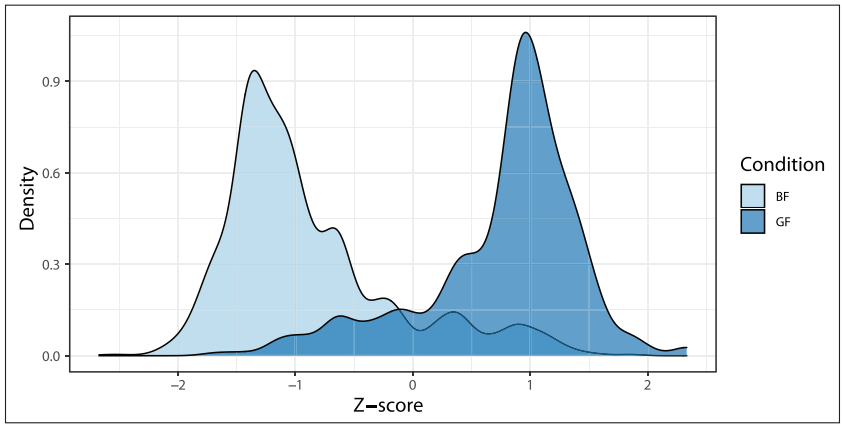


Figure 8: Distribution of z-scores for the fillers, Experiment 2. The bad fillers show a narrow distribution around –1.5. The good fillers show a narrow distribution around 1.

[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.

	Number of data points	***Estimate***	***t***	***p***
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.

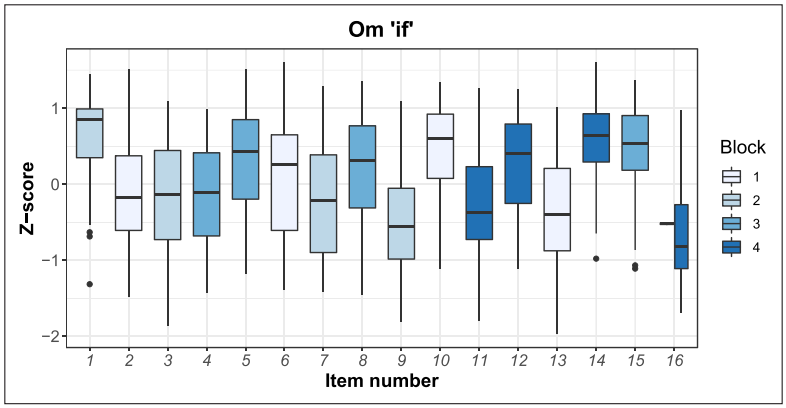


Figure 9: Boxplot illustrating the average judgments on the "long, island" condition by item. The different shades of blue indicate the different block numbers.

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.

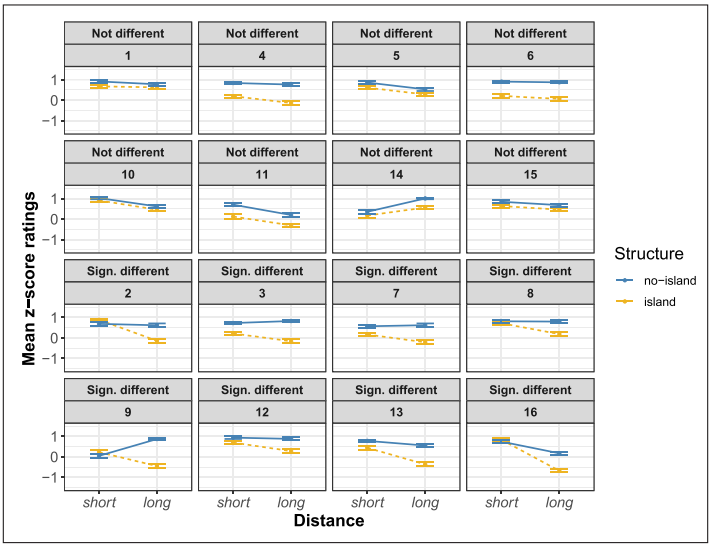


Figure 10: Interaction plot by item for *om* 'if', Experiment 2. The items that the model did not distinguish from item 1 are labelled "Not different", while the items that the model did distinguish from item 1, "Sign. different".

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 aggregrated 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.

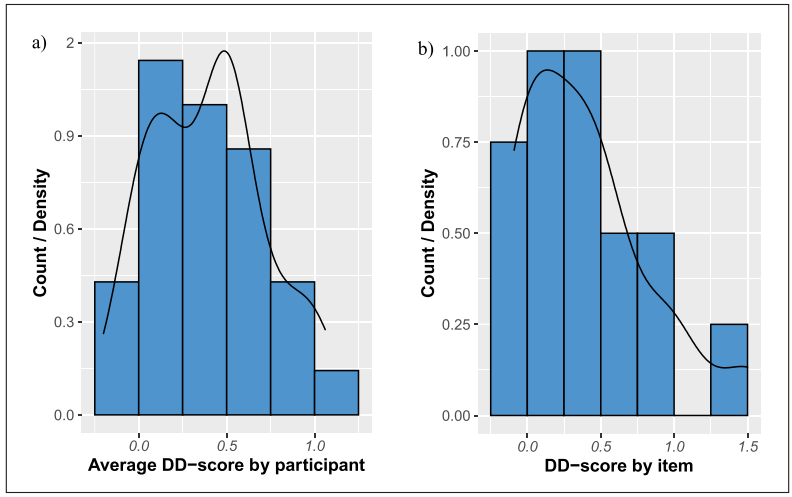


Figure 11: DD-scores calculated by participant across 16 items (a) and DD-scores calculated by item across 56 participants (b), Experiment 2. Histograms are plotted with geom_histogram(), boundary = 0, binwidt = 0.25.[18]

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 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.

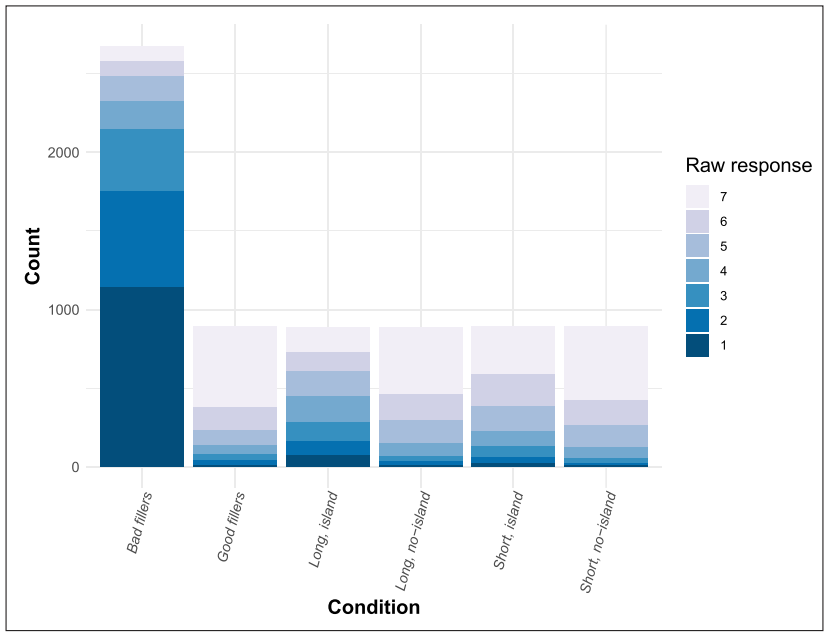


Figure 12: Barplots displaying the count of raw responses per condition, Experiment 2.

[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 kunstsamleren som jubler om / når / fordi noen*
Nils talks with art dealer.DEF who celebrates if / when / because someone
kjøper maleriet av Van Gogh.[22]
buys paitning.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 decribed 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

Åfarli, 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 Portugese 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 Univserity 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. *Procedings 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 sensitvity of accetpability 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

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 & Calcagnì. 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 Linguistic* 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 licensor 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 Satser 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.

Paper 3

ISBN 978-82-326-7914-0 (printed ver.)
ISBN 978-82-326-7913-3 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology