



Fukuda, Shin & Tanaka, Nozomi & Ono, Hajime & Sprouse, Jon. 2022. An experimental reassessment of complex NP islands with NP-scrambling in Japanese. *Glossa: a journal of general linguistics* 7(1), pp. 1–35. DOI: <https://doi.org/10.16995/glossa.5737>



Open Library of Humanities

An experimental reassessment of complex NP islands with NP-scrambling in Japanese

Shin Fukuda, Department of Linguistics, University of Hawai'i at Mānoa, fukudash@hawaii.edu

Nozomi Tanaka, Department of East Asian Languages and Cultures, Indiana University, tanakan@iu.edu

Hajime Ono, Department of English, Tsuda University, hajime@tsuda.ac.jp

Jon Sprouse, Program in Psychology, New York University Abu Dhabi, jon.sprouse@nyu.edu

There is little consensus in the Japanese syntax literature on the question of whether complex NPs with a noun complement headed by *toyuu* 'that.say' are islands for NP-scrambling dependencies. To explore this question, we conducted two acceptability judgment experiments using the factorial definition of islands to test the status of noun complements, relative clauses (which are also complex NPs, and uniformly considered islands in the literature), and coordinated NP structures (which are also uniformly considered islands in the literature). Our first experiment yielded strong evidence that relative clauses and coordinated NPs are islands (as expected), and strong evidence that noun complements are not. Our second experiment also found strong evidence that relative clauses and coordinated NPs are islands, but yielded a small, non-significant, trend toward an effect with noun complements. Based on the sizes of our samples (89 and 90 participants, respectively), the sizes of the effects, and the details of the acceptability patterns, we conclude that noun complements in Japanese are not islands with respect to NP-scrambling. We also investigated between- and within-participant variability in our results. We observe no evidence of increased between-participant variability for noun complements relative to other islands, and no increase of within-participant variability for noun complements relative to scrambling out of (non-island) declarative CPs. Our results have consequences for a number of issues that have been encoded in current syntactic theories of island effects, including the correlation between syntactic constituent complexity and island status (e.g., number of bounding nodes or phase heads), and the correlation between complementizer deletion and island status (e.g., the complement/adjunct distinction).

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1 Introduction

There is little consensus in the Japanese syntax literature on the question of whether complex NPs with a noun complement headed by *toyuu* ‘that.say’ (henceforth, noun complements) are islands for NP-scrambling dependencies. For example, Haig (1976) claims that noun complements are not islands for NP-scrambling, whereas relative clauses are islands for NP-scrambling. In contrast, Saito (1985) claims that both noun complements and relative clauses are islands for NP-scrambling, but that the island effect of noun complements is smaller than the island effect of relative clauses. Recent experimental work has added to this uncertainty. Yano (2019), as part of a broader study of the effect of D-linking on islands in Japanese, tested noun complements (but not relative clauses) in two acceptability judgment experiments using the factorial definition of island effects. However, the two experiments investigating noun complements with non-D-linked NPs produced potentially conflicting results: the first experiment revealed a (relatively small) island effect, whereas the second experiment revealed no island effect (see Section 2 for additional discussion). This suggests a need for additional systematic data collection. Therefore, in this study, we present two additional judgment studies specifically designed to explore complex NP islands with NP-scrambling in Japanese (Experiments 1 and 2). We use the factorial definition of island effects to explore the status of NP-scrambling out of both noun complements and relative clauses, and for additional comparison of the size of the island effects (following Saito’s 1985 suggestion), we also include NP-scrambling out of coordinated NP structures (henceforth, coordinate structures), which is uncontroversially considered an island for NP-scrambling in Japanese (Harada 1977).

The rest of this paper is organized as follows. Section 2 provides some background on noun complements and NP-scrambling in Japanese, including a recent experimental study (Yano 2019) which tested NP-scrambling out of noun complements, as well as the general motivations for our study. Section 3 reports the design and results of our first experiment. In that experiment, we attempted to use the typical factorial design for island effects and collect a relatively large sample (89 participants) to establish the status of noun complements. Anticipating our results slightly, Experiment 1 yields strong evidence that relative clauses and coordinate structures are islands, but noun complements are not. We also find no evidence of differences in between-participant variation among the three island types. Section 4 reports the design and results of our second experiment. In that experiment, we used a novel factorial design to investigate the generalizability of our results and to investigate both between- and within-participant variation. The results again suggest strong evidence that relative clauses and coordinate structures are islands; however, the results for noun complements are more complicated, as the two statistical tests yield contradictory results. Based on the overall pattern of the results in Experiment 2, we conclude that noun complements are, at the very least, not classic island effects, and consequently substantially different than relative clauses and coordinate structures. Thus, taken as a whole, the results of our two experiments show that there is a clear difference between

relative clauses and noun complements, at least in Japanese, and more broadly, between noun complements in Japanese and noun complements in other languages that have been tested using the factorial definition, such as English, Italian, and Norwegian (cf. Sprouse et al. 2011; Sprouse et al. 2012; Sprouse et al. 2016; Kush et al. 2018). Section 5 provides a discussion of the theoretical consequences of our results. We argue that our findings have direct consequences for most existing theories of island effects, for theories of the relationship between complementizer deletion and island status, and for theories of the relationship between the complexity of syntactic structure and island status. We further suggest that future studies should probe the properties of relative clauses and noun complements (cross-linguistically) along these dimensions.

2 Complex NPs and NP-scrambling in Japanese

In this section we provide a brief description of two types of complex NPs in Japanese that are the topic of this study and review previous claims about complex NP islands with NP-scrambling in Japanese.

2.1 Noun complements and relative clauses in Japanese

As the main empirical goal of our study is to compare scrambling out of noun complements and relative clauses in Japanese, some discussion of the syntactic properties of noun complements and relative clauses in Japanese is in order.

Noun complements in this study are complex NPs headed by *toyuu* ‘that.say’ and nouns such as *uwasa* ‘rumor’ and *shooko* ‘evidence’ as in (1a–b) below.

- (1) a. Taro-wa [_{NC} Jiro-ga ki-ta toyuu uwasa]-o kii-ta.
 T-TOP J-NOM come-PST that.say rumor-ACC hear-PST
 ‘Taro heard the rumor that Jiro came.’
- b. Taro-wa [_{NC} Jiro-ga ki-ta toyuu shooko]-o mitsuke-ta.
 T-TOP J-NOM come-PST that.say evidence-ACC find-PST
 ‘Taro found the evidence that Jiro came.’

Following previous studies (e.g., Fukui 1988), we assume that *toyuu* ‘that.say’ is a complementizer and that noun complements are CPs. There are at least two other types of complex NPs in Japanese: those that are headed by *no* (2a) and those that are headed by *koto* (2b).

- (2) a. Taro-wa [_{CNP} Jiro-ga ki-ta no]-o mi-ta.
 T-TOP J-NOM come-PST NO-ACC see-PST
 ‘Taro saw Jiro come.’
- b. Taro-wa [_{CNP} Jiro-ga ki-ta koto]-o shit-ta.
 T-TOP J-NOM come-PST KOTO-ACC know-PST
 ‘Taro found out that Jiro came.’

This study focuses on *toyuu* complex NPs as in (1) for the following reasons. First, as discussed in the introduction, previous studies on scrambling out of complex NPs in Japanese focused on *toyuu* complex NPs presumably because they represent the Japanese equivalents of noun complements in English (e.g., Nakau 1973). Second, although all three types of complex NPs in (1) and (2) exhibit the basic syntactic properties of NPs, such as being marked by a case marker and functioning as subjects and direct objects, the nouns like *uwasa* ‘rumor’ and *shooko* ‘evidence’ in (1) are different from *no* and *koto* in (2) as the former have a clear and identifiable meaning while the latter are semantically bleached.¹ There are also indications that *no* and *koto* have become “markers” of certain semantic/pragmatic distinctions beyond the typical functional role of nouns. According to Kuno (1973), the semantic/pragmatic contribution of *no* and *koto* is that the complex NPs headed by them represent propositions that the speaker presupposes to be true. Relatedly, nouns like *uwasa* ‘rumor’ can stand alone as full NPs, whereas *no* and *koto* cannot. Under the assumption that noun complements involve the head noun taking a clausal complement, the head noun in a noun complement must be a lexical item with the ability to thematically license its complement. Lexical nouns like *uwasa* ‘rumor’ and *shooko* ‘evidence’ fit this description, while *no* and *koto* do not.

Relative clauses and noun complements in Japanese differ in the same ways that relative clauses and noun complements differ in many other languages (including English): for example, noun complements stand in a thematic relation with the selecting noun, while relative clauses are modifiers; similarly, in noun complements every argument required by the thematic structure of the embedded verb is present, while relative clauses have a missing argument that is interpreted as coreferential with the head NP (the relative clauses used in this study do not allow resumptive pronouns, but see Kuno 1973 for the claim that certain relative clauses, such as those with multiple levels of embedding may allow resumptive pronouns). However, noun complements and relative clauses also differ in a way that is not, to our knowledge, as common: at least two pieces of evidence suggest that the structure of noun complements in Japanese is more complex along certain dimensions than the structure of the relative clauses. First, Tomioka (2015) observes that noun complements can embed an NP marked by *-wa* (3a) while relative clauses cannot (3b), potentially suggesting an additional layer of functional structure in noun complements under the assumption that topics appear higher in the left periphery than nominative subjects.

- (3) a. [_{NC} Erika-ga/wa kekkon shi-ta toyuu uwasa]
 E-NOM/TOP marriage do-PST that.say rumor
 ‘the rumor that Erika got married’
- b. [_{RC} Erika-ga/*wa ec_i kat-ta kuruma_i]
 E-NOM/*TOP buy-PST car
 ‘a/the car Erika bought’

¹ For instance, Josephs (1976) analyzes these items as nominalizers.

Second, it has also been observed that relative clauses allow a non-episodic interpretation of verbs that normally entails a change of state, as in (4a) (e.g., Teramura 1982; Ogihara 2004). Noun complements do not allow such an interpretation of similar verbs (4b).

- (4) a. Taro-wa [_{RC} kawai-ta taoru]-o tanon-da.
 T-TOP dry-PST towel-ACC ask-PST
 ‘Taro asked for a dry towel/a dried towel.’
- b. Taro-wa [_{NC} taoru-ga kawai-ta toyuu hookoku]-o shi-ta.
 T-TOP towel-NOM dry-PST that.say report-ACC do-PST
 ‘Taro made a report that the towel dried.’

(4a) is ambiguous between two interpretations. One interpretation is that what Taro asked for is a dry towel, while the alternative interpretation is that he asked for a dried towel, a towel that was wet at some point in the past, but it had dried. In (4b), in contrast, the embedded sentence can only have the latter interpretation, i.e., a towel underwent a change of state from ‘not dry’ to ‘dry.’ Ogihara (2004) argues that relative clauses such as the one in (4a) do not involve a clausal structure but a TP with a reduced verbal projection.

In our study, noun complements always involve a noun preceded by an embedded clause with a full set of arguments marked by *toyuu* ‘that.say,’ while relative clauses always involve a gap that is identified with the modified NP and are never marked by *toyuu* ‘that.say.’²

2.2 The island status of noun complements with NP-scrambling

Haig (1976) was one of the first theoretical studies to investigate complex NPs in Japanese, reporting that NP-scrambling out of a noun complement is acceptable (5a), while NP-scrambling out of a relative clause is not (5b). The judgments in (5) are from Haig (1976).

² *Toyuu* ‘that.say’ can appear in relative clauses, so it is not, by itself, an unambiguous signal of noun complements:

- (i) [_{RC} Erika-ga ec_i kat-ta toyuu kuruma_i]
 E-NOM buy-PST that.say car
 ‘a/the car Erika bought’

However, as Nakau (1973) observed, relative clauses without a gap in an argument position cannot be followed by *toyuu* ‘that.say.’ Therefore, the absence of a gap and the presence of *toyuu* is an unambiguous signal:

- (ii) a. [_{RC} dareka-ga samma-o yak-u nioi]
 someone-NOM mackerel-ACC broil-NPST smell
 ‘the smell of someone broiling a mackerel.’
- b. * [_{RC} dareka-ga samma-o yak-u toyuu nioi]
 someone-NOM mackerel-ACC broil-NPST that.say smell
 (‘the smell of someone broiling a mackerel.’)

- (5) a. Mary-o_i watashi_k-wa [_{NC} ec_k Bill-ni t_i shookaishi-ta-i toyuu kiboo]-o
 M-ACC I-TOP B-DAT introduce-want-NPST that.say desire-ACC
 mottei-ru.
 have-NPST
 ‘I have a desire such that I want to introduce Mary to Bill.’ (Haig 1976: 369; (25))
- b. *Ano hon-o_i watashi-wa [_{RC} ec_j t_i kai-ta hito_j]-ni ai-ta-i.
 that book-ACC I-TOP write-PST person-to meet-want-NPST
 ‘I want to meet the person who wrote that book.’ (Haig 1976: 370; (30))

Saito (1985; 1987) made the more nuanced claim that both noun complements and relative clauses are islands, with noun complements being relatively more acceptable (6a) than relative clauses (6b). The judgments in (6) are from Saito (1985).

- (6) a. ?Bill-o_i John-ga [_{NC} Mary-ga t_i saketei-ru toyuu uwasa]-o kii-ta.
 B-ACC J-NOM M-NOM avoid-NPST that.say rumor-ACC hear-PST
 ‘John heard a rumor that Mary is avoiding Bill.’ (Saito 1985: 246; (146b))
- b. ?*Ano hon-o_i John-ga [_{RC} ec_j t_i kat-ta hito_j]-o sagashitei-ru rasii.
 that book-ACC J-NOM buy-PST person-ACC search-NPST seem
 ‘It seems that John is looking for the person who bought that book.’
 (Saito 1985: 246; (146a))

To the best of our knowledge, Yano (2019) is the only published study to examine the acceptability of NP-scrambling out of noun complements with formal acceptability judgment experiments. The goal of Yano (2019) was to examine whether D-linked NPs like *sono shoosetsu* ‘the novel’ undergo syntactic movement when they appear in a fronted position. Yano (2019) uses island effects as a diagnostic of movement. To that end, Yano tested two island types: adjunct islands and noun complement islands. Yano (2019) tested both D-linked NPs (with *sono* ‘the/its’) as the target of investigation, and non-D-linked NPs (without *sono* ‘the/its’) as a baseline comparison. Here we focus exclusively on non-D-linked NPs because the effect of D-linked phrases, or lack thereof, is a potentially more complex topic of investigation that takes the island facts for non-D-linked NPs as a starting point (see Szabolcsi & Lohndal 2017 for a review of selective islands). In what follows, we focus on the discussion in Yano (2019) of noun complement islands.

In the first experiment of Yano (2019), the sentences were presented in isolation. In the second experiment, the sentences were presented with a context sentence such as “The novel received the Naoki prize.” to establish the fronted object in the discourse.³ Yano (2019) used the factorial definition of island effects in which the presence of an island effect appears as a superadditive interaction of two (or more) factors that are themselves independent of the

³ The Naoki prize is a prestigious Japanese literary award.

island effects (Sprouse 2007; Sprouse et al. 2011; Sprouse et al. 2012, a.o.). For the Yano (2019) experiments these factors were STRUCTURE, manipulating the structure of the embedded clause (either an island or a non-island), and WORD ORDER, manipulating the presence or absence of scrambling out of the embedded clause. An example set of the four conditions is given in (7) for completeness; we review the logic of the factorial design, and provide full examples for our experiments, in Section 3.

- (7) Example experimental sentences from Yano (2019: 5, ex.9, gloss modified)
- a. non-island / no-scrambling
 Hyooronka-wa [_{CP} kyonen goosutoraitaa-ga (sono) shoosetsu-o
 commentator-TOP last.year ghost.writer-NOM (the) novel-ACC
 kai-ta-to] shinjitei-ru.
 write-PST-COMP believe-NPST
 ‘The commentator believes that the ghost-writer wrote (the) novel last year.’
 - b. non-island / scrambling
 (Sono) shoosetsu-o_i hyooronka-wa [_{CP} kyonen goosutoraitaa-ga *t*_i
 (the) novel-ACC commentator-TOP last.year ghost.writer-NOM
 kai-ta-to] shinjitei-ru.
 write-PST-COMP believe-NPST
 - c. island / no-scrambling
 Hyooronka-wa [_{NC} kyonen goosutoraitaa-ga (sono)shoosetsu-o_i kai-ta
 commentator-TOP last.year ghost.writer-NOM (the) novel-ACC write-PST
 toyuu hoodoo-o] shinjitei-ru.
 that.say news-ACC believe-NPST
 ‘The commentator believes the news that the ghost-writer wrote (the) novel last year.’
 - d. island / scrambling
 (Sono) shoosetsu-o_i hyooronka-wa [_{NC} kyonen goosutoraitaa-ga *t*_i kai-ta
 (the) novel-ACC commentator-TOP last.year ghost.writer-NOM write-PST
 toyuu hoodoo-o] shinjitei-ru.
 that.say news-ACC believe-NPST

The results of the two Yano (2019) experiments are potentially conflicting. In the first experiment (without context), Yano (2019) finds a small superadditive interaction indicative of a noun complement island effect. In the second experiment (with context), Yano (2019) finds no superadditive interaction indicative of a noun complement island effect. There are at least two issues that make the Yano (2019) results difficult to interpret with respect to the question of the islandhood of noun complements (which, to be fair, was not the research question for Yano 2019). First, the Yano (2019) results showed very low acceptability even for putatively grammatical scrambling out of non-island embedded clauses (a declarative CP), in both experiments. Yano

notes that this could be due to the *long-before-short* preference – that is, a preference in Japanese that scrambled NPs be longer than the NPs that they are scrambled over (Dryer 1980; Hawkins 1994; Yamashita & Chang 2001; Yamashita 2002; Omaki et al. 2020). The scrambled NPs in the Yano 2019 experiments are single-word NPs (presumably because of the focus of the study on D-linking), which violates the long-before-short preference, and therefore could have pushed the overall acceptability down. This in turn could have reduced the size of the superadditive interactions (if the long-before-short preference is not additive with island effects, which is itself a potentially interesting observation that might merit future study). Second, as mentioned above, the second experiment manipulated the presence versus absence of a context. The results show that there was no superadditive interaction with the “without context” (“non-D-linked” in Yano’s terms) sentences, while there was a small superadditive interaction with the “with context” (“D-linked” in Yano’s terms) sentences. Here, it is important to reiterate that the experimental sentences for the “without context” condition in the second experiment were identical with respect to island-relevant properties to the non-D-linked experimental sentences in the first experiment (e.g., the scrambled NPs were bare NPs without a demonstrative). Thus, the superadditive interaction observed in the first experiment disappeared in the second experiment despite the fact that virtually identical sentences were judged in these two experiments. Furthermore, the directionality of the result – no island effect without context and an island effect with context – runs contrary to the directionality predicted by theories that predict D-linking to ameliorate island effects. Given the complex pattern of results between the two experiments, Yano (2019) speculates that noun complements may show more between-participant variability than other island types.

2.3 The motivation of the current study

The contradictory results for noun complements between Haig (1976) and Saito (1985; 1987), and between the two experiments in Yano (2019), suggest that additional systematic data collection is needed, with special attention paid to the long-before-short preference as well as to the possibility that noun complements show more between-participant variability than other island types. To that end, here we report the results of two formal acceptability judgment experiments testing whether NP-scrambling out of complex NPs invokes island effects, both with materials that respect the long-before-short preference, and both with large sample sizes (89 and 90 participants, respectively) to allow for high statistical power and the possibility of exploring between-participant variability.⁴

⁴ The two experiments we report in this paper were first conducted in 2017 as part of a larger project investigating island structures in Japanese, and the original manuscript submitted to *Glossa* was based on the data from these experiments. After the first round of peer review in 2021, we recruited additional participants for the same experiments in order to improve the statistical power of our analyses. As such, our two experiments predate the experi-

Scrambling dependencies are conventionally used to test island effects in Japanese due to the lack of overt *wh*-movement (for a discussion of island effects involving *wh-in-situ*, see Sprouse et al. 2011; Kim & Goodall 2016; Tanaka & Schwartz 2018; Lu et al. 2020). But before we move on to discuss our experiments, a caveat is in order concerning some characteristics of scrambling and the design of our experiments. First, it has been argued that some instances of scrambling exhibit properties of A-bar-movement while others exhibit properties of A-movement. In our study, all instances of scrambling are long distance, i.e., they always cross a clausal boundary. Since the consensus in the literature is that long distance scrambling is A-bar-movement (e.g., Saito 1992; Yoshimura 1992; Nemoto 1993; Tada 1993; see Nemoto 1999 for a comprehensive review of the relevant literature), we assume that all the instances of scrambling examined in this study are instances of A-bar-movement. Second, unlike *wh*-movement and relativization, scrambling is optional and has been claimed to be “semantically vacuous” (e.g., Saito 1989; but see Miyagawa 2001 for a claim that local scrambling can be triggered by the EPP and, therefore, obligatory). The fact that scrambling is an optional process raises questions about its motivations. Factors such as constituent weight (Yamashita & Chang 2001; Omaki et al. 2020) and information-structure status (Koizumi & Imamura 2017) have been shown to affect the production, acceptability, and processing of scrambling sentences. The optional nature of scrambling also raises the possibility that the effect of scrambling on acceptability judgments in non-island environments might be different from the effects of *wh*-movement and relativization on acceptability judgments in similar environments. Though the question of whether scrambling differs qualitatively and/or quantitatively from the other types of A-bar dependencies is interesting in its own right, we believe it is beyond the scope of our study, as it requires a broad set of experiments that compares different types of A-bar dependencies, whereas our experiments were narrowly constructed to explore NP-scrambling. Finally, there is one important difference between our experiments with NP-scrambling and previous studies that investigated *wh*-questions: whereas previous studies of *wh*-questions manipulated the distance of *wh*-movement dependencies, with *wh*-movement originating in either the matrix clause (short) or embedded clause (long), our experiments on NP-scrambling (and those in Yano 2019) manipulated the presence or absence of long-distance scrambling, not the distance of the scrambling. This is because an instance of NP-scrambling is unambiguously A-bar-movement only if it is long distance – thus, one cannot compare instances of short and long scrambling sentences without introducing yet another factor, such as the A-bar vs A distinction, that might affect their acceptability. We therefore note that the presence of scrambling may incur a larger main effect, as the mere presence of a long-distance dependency alone has been shown to cause a

ments in Yano (2019), and were not (and could not be) designed specifically to clarify or extend those results. Nonetheless, we are indebted to Yano (2019) for further revealing the uncertainty surrounding the islandhood of noun complements as part of his D-linking study.

significant decrease in acceptability compared to sentences without a long-distance dependency (e.g., Kluender & Kutas 1993). A potentially larger main effect in turn raises the possibility that the superadditive interaction might cause a floor effect; we discuss this possibility as part of the description of the logic of the design in Section 3.1 below, and we note that there is no evidence of floor effects in our results in Sections 3.4 and 4.4.

3 Experiment 1

3.1 The logic of the design

Experiment 1 tested three island types: noun complements, relative clauses, and coordinate structures. By including both types of complex NPs together in the same experiment, we can investigate Saito's (1985; 1987) claim that noun complements yield smaller island effects than relative clauses. We included coordinate structures because they are uncontroversially considered islands in the literature (Harada 1977), and therefore can serve as a type of baseline comparison for both relative clause islands and noun complements.

We employed the factorial definition of island effects, both because we believe it matches the logic that has historically been used by syntacticians to define island effects, and because it allows us to eventually integrate our results with the growing cross-linguistic experimental literature using the factorial definition (a.o., Christensen et al. 2013;⁵ Almeida 2014; Kim & Goodall 2016; Sprouse et al. 2016; Keshev & Meltzer-Asscher 2018; Kush et al. 2018; 2019; Stepanov et al. 2018; Tanaka & Schwartz 2018; Ko et al. 2019; Lu et al. 2019; Tucker et al. 2019; Omaki et al. 2020). As described below, we implement the factorial design completely within participants, allowing us to quantify to what extent each participant reports an island effect, so that we can investigate the conjecture from Yano (2019) that noun complements may show a higher degree of between-participant variability than other island types. Finally, we use relatively long scrambled NPs to satisfy the long-before-short preference.

The factorial design has two factors: SCRAMBLING manipulates the presence or absence of NP-scrambling (no-scrambling/scrambling), and STRUCTURE manipulates the structure of the embedded clause (non-island/island). Fully crossing these two factors in a 2×2 design leads to four conditions. In (8), we illustrate all four conditions for noun complements. Note that the NPs that are the target of scrambling are outlined with a box.⁶

⁵ The design and analysis of the experiments in Christensen et al. 2013 are not specifically intended to follow the factorial logic we use here, but we include that study here because the four conditions necessary for the factorial logic are present in the design.

⁶ The complete lists of the experimental and filler sentences used in Experiment 1 can be found in the OSF repository for the current study (DOI [10.17605/osf.io/w2zdvdv](https://doi.org/10.17605/osf.io/w2zdvdv)).

(8) Example conditions for noun complements

a. non-island / no-scrambling

Shinjin-no kisha-wa [_{CP} IT-gaisha-no shachoo-ga chuumoku-no
 novice-GEN reporter-TOP IT-company-GEN CEO-NOM popular-GEN
kakkitekina sofutowea-o daigaku.zaigakuchuu-ni kaihatsushi-ta-to]
 innovative software-ACC college.days-in develop-PST-COMP
 kiji-ni kai-ta.
 article-as write-PST

‘That novice reporter wrote (as an article) that the CEO of the IT company developed the popular, innovative software while s/he was in college.’

b. non-island / scrambling

Chuumoku-no kakkitekina sofutowea-o_i shinjin-no kisha-wa [_{CP}
 popular-GEN innovative software-ACC novice-GEN reporter-TOP
 IT-gaisha-no shachoo-ga t_i daigaku.zaigakuchuu-ni kaihatsushi-ta-to]
 IT-company-GEN CEO-NOM college.days-in develop-PST-COMP
 kiji-ni kai-ta.
 article-as write-PST

c. island / no-scrambling

Shinjin-no kisha-wa [_{NC} IT-gaisha-no shachoo-ga chuumoku-no
 novice-GEN reporter-TOP IT-company-GEN CEO-NOM popular-GEN
kakkitekina sofutowea-o daigaku.zaigakuchuu-ni kaihatsushi-ta-toyuu
 innovative software-ACC college.days-in develop-PST-that.say
 kiji-o] kai-ta.
 article-ACC write-PST

‘That novice reporter wrote an article that the CEO of the IT company developed the popular, innovative software while s/he was in college.’

d. island / scrambling

Chuumoku-no kakkitekina sofutowea-o_i shinjin-no kisha-wa [_{NC}
 popular-GEN innovative software-ACC novice-GEN reporter-TOP
 IT-gaisha-no shachoo-ga t_i daigaku.zaigakuchuu-ni kaihatsushi-ta-toyuu
 IT-company-GEN CEO-NOM college.days-in develop-PST-that.say
 kiji-o] kai-ta.
 article-ACC write-PST

(9) and (10) below show all four conditions for relative clauses and coordinate structures, respectively. The non-island structure that we chose for relative clauses was a declarative CP. The non-island structure that we chose for coordinate structures was an NP-PP sequence.

(9) Example conditions for relative clauses

a. non-island / no-scrambling

Roodookumiai-no riidaa-wa [_{CP} kaisha-no juuyaku-ga oohabana
 union-GEN leader-TOP company-GEN executives-NOM drastic
uriage-no nobi-o juugyooiin-no kyuuuyo-ni han'ee saseteinai-to]
 sales-GEN growth-ACC employee-GEN salary-to reflect not.make-COMP
 hihanshi-ta.
 criticize-PST

'The union leader criticized that the executives of the company were not making the drastic sales growth reflected in the employees' salaries.'

b. non-island / scrambling

Oohabana uriage-no nobi-o_i roodookumiai-no riidaa-wa [_{CP} kaisha-no
 drastic sales-GEN growth-ACC union-GEN leader-TOP company-GEN
 juuyaku-ga t_i juugyooiin-no kyuuuyo-ni han'ee saseteinai-to] hihanshi-ta.
 executives-NOM employee-GEN salary-to reflect not.make-COMP criticize-PST

c. island / no-scrambling

Roodookumiai-no riidaa-wa [_{RC} ec_i oohabana uriage-no nobi-o juugyooiin-no
 union-GEN leader-TOP drastic sales-GEN growth-ACC employee-GEN
 kyuuuyo-ni han'ee saseteinai kaisha-no juuyaku_i-o] hihanshi-ta.
 salary-to reflect not.make company-GEN executives-ACC criticize-PST

'The union leader criticized the company's executives who were not making the drastic sales growth reflected in the employees' salaries.'

d. island / scrambling

Oohabana uriage-no nobi-o_i roodookumiai-no riidaa-wa [_{RC} $ec_j t_i$
 drastic sales-GEN growth-ACC union-GEN leader-TOP
 juugyooiin-no kyuuuyo-ni han'ee saseteinai kaisha-no juuyaku_j-o]
 employee-GEN salary-to reflect not.make company-GEN executives-ACC
 hihanshi-ta.
 criticize-PST

(10) Example conditions for coordinate structures

a. non-island / no-scrambling

Sono kaishain-wa tenkinsu-ru buka-ga [_{NP} kompooshi-ta hikkoshi-no
 that employee-TOP transfer-NPST subordinate-NOM boxed-PST moving-GEN
nimotsu-o] [_{PP} kattabakari-no shinsha-to] tomoni fune-de hikkoshisaki-ni
 items-ACC just.purchased-GEN new.car-with together ship-by new.address-to
 okut-ta-to setsumeeshi-ta.
 send-PST-COMP explain-PST

'The employee explained that the subordinate who was being transferred shipped boxes of items together with a new car that s/he just purchased to the new address.'

b. non-island / scrambling

[_{NP} Kompooshi-ta hikkoshi-no nimotsu-o_i] sono kaishain-wa tenkinsu-ru
 box-PST moving-GEN items-ACC that employee-TOP transfer-NPST
 buka-ga t_i [_{PP} kattabakari-no shinsha-to] tomoni fune-de
 subordinate-NOM just.purchased-GEN new.car-with together ship-by
 hikkoshisaki-ni okut-ta-to setsumeeshi-ta.
 new.address-to send-PST-COMP explain-PST

c. island / no-scrambling

Sono kaishain-wa tenkinsu-ru buka-ga [_{NP} kompooshi-ta]
 that employee-TOP transfer-NPST subordinate-NOM boxed-PST
 hikkoshi-no nimotsu-to kattabakari-no shinsha-o] tomoni fune-de
 moving-GEN items-and just.purchased-GEN new.car-ACC together ship-by
 hikkoshisaki-ni okut-ta-to setsumeeshi-ta.
 new.address-to send-PST-COMP explain-PST
 ‘The employee explained that the subordinate who was being transferred shipped
 boxes of items and a new car that s/he just purchased together to the new address.’

d. island / scrambling

[Kompooshi-ta hikkoshi-no nimotsu-to_i] sono kaishain-wa tenkinsu-ru
 boxed-PST moving-GEN items-and that employee-TOP transfer-NPST
 buka-ga [_{NP} t_i kattabakari-no shinsha-o] tomoni fune-de
 subordinate-NOM just.purchased-GEN new.car-ACC together ship-by
 hikkoshisaki-ni okut-ta-to setsumeeshi-ta.
 new.address-to send-PST-COMP explain-PST

The value of the factorial definition is that it isolates the island effect in the interaction between SCRAMBLING and STRUCTURE (while subtracting out the main effects of those factors). If there is no island effect, we expect to see no interaction as illustrated in the left panel of **Figure 1**, where the two lines that connect the two means for the island condition sentences and the non-island condition sentences are parallel. If there is an island effect, we expect to see a superadditive interaction as illustrated in the center and right panels, where the two lines are not parallel because the mean for the scrambling/island condition sentence is lower than expected if the effects of the two manipulations are all there are. Crucially, we can also look at the size of the interaction as a measure of the size of the island effect (e.g., to test the claim by Saito 1985; 1987); the center panel illustrates a smaller effect, and the right panel illustrates a larger effect.

As discussed in Section 2.2, one factor that makes our experiments different from previous studies that examined other types of A-bar dependencies is that our second factor is the absence versus presence of an A-bar dependency, i.e., NP-scrambling, while previous studies manipulated the distance of the dependency (e.g., *wh*-movement that originated in the matrix versus embedded clause). Because of this difference, our experiments might show a larger main effect of the second

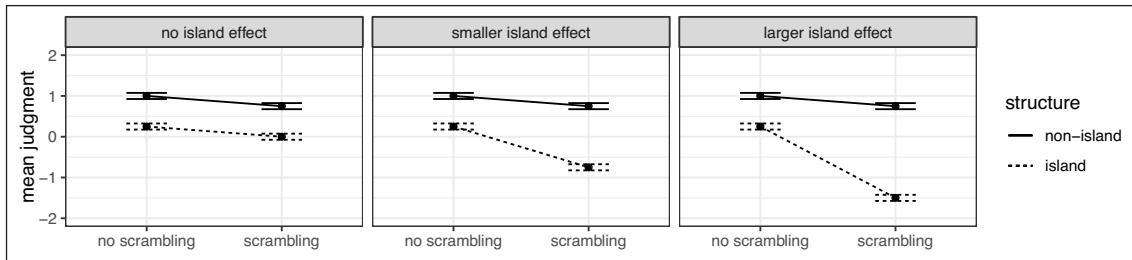


Figure 1: Possible results using the factorial definition of island effects.

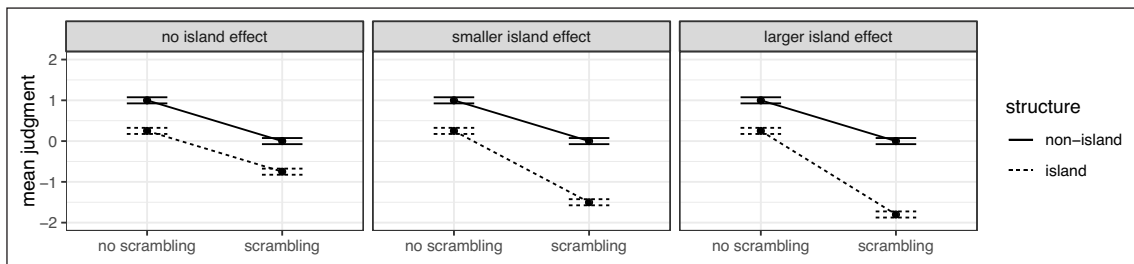


Figure 2: Possible results with hypothesized larger main effects of scrambling.

factor than is observed in previous studies. **Figure 2** demonstrates this. The larger main effect than in **Figure 1** appears as a steeper downward slope in the non-island structure line.

One concern that arises with large main effects is that they make a floor effect likely with superadditive interaction terms. A floor effect arises when the superadditive interaction is so large that it should push the island/scrambling condition beneath the lower bound of the scale, but because the scale does not go any lower, the island/scrambling condition is rated higher than it should be (because it is metaphorically stopped by the floor of the scale). The end result is an underestimation of the island effect size. Though we cannot eliminate this possibility, we can check for the possibility of floor effects by plotting the rating of the least acceptable filler in each plot as a solid gray line as an estimate of the functional floor of the scale. If the island/scrambling condition is lying precisely on this line, then a floor effect is likely (though not certain). If the island/scrambling condition is above this line, then a floor effect is not likely.

3.2 Materials and survey construction

Each participant completed a survey that consisted of 58 items: 6 practice items, 12 experimental items and 40 filler items pseudorandomized to avoid related experimental items appearing in succession. The 12 experimental items consisted of 1 token of each of the 4 conditions for each of the three islands. We chose one judgment per condition per participant to keep the total number of experimental items low to minimize the chance that participants would notice the goal of the experiment. We compensated for the increased risk of noise with one judgment per

Condition	Noun Complement	Relative Clause	Coordinate Structure
Non-island/no-scrambling	89	100	89
Non-island/scrambling	89	95	85
Island/no-scrambling	89	83	93
Island/scrambling	89	78	89

Table 1: The number of observations per condition (per island) in Experiment 1.

condition by testing a sample size ($n = 89$) that is likely to yield very high statistical power for medium and large effect sizes, and moderate statistical power for small effect sizes (Sprouse & Almeida 2017). We created 8 lexically matched sets (of 4 conditions) of items per island. The items were then distributed among 8 experimental lists, each 4 items long (one per condition in the factorial design), using a Latin square procedure so that participants saw a unique lexical item in each condition. We identified 4 errors in the item codes (out of 96 items across lists) after the experiment. We corrected these errors during analysis, but it meant that the total number of observations per condition were mildly uneven (Table 1).

3.3 Participants and presentation

Ninety-one participants from two universities in Tokyo, Japan, participated in the experiment. We excluded two participants from analysis because their answers to our language background questionnaire revealed that they had significant exposure to a language other than Japanese before they were 10-years old. Eighty-nine self-reported native speakers of Japanese remain in the analysis. Participants either received course credit for their participation or 500 yen. The experiment was administered online using IBEX (Drummond 2013). Each sentence was presented one at a time on its own presentation screen with a 1 (*mattaku fushizen* ‘completely unnatural’) to 7 (*mattaku shizen* ‘completely natural’) scale. Participants indicated their rating by clicking on the appropriate number. Because complex NP islands may show variability across participants, we did not exclude any participants from analysis because of the distribution of their judgments.

3.4 Results

In this section we describe the results of Experiment 1, with a particular focus on (i) the presence or absence of the superadditive interaction indicative of island effects (and, relatedly, the relative size of the effect) and (ii) the variability of island effects across participants.

3.4.1 The presence or absence of island effects

To determine the presence or absence of island effects, we will look for two properties: (i) a visual pattern indicating a superadditive interaction among the 4 conditions in the factorial design (as illustrated in **Figures 1 and 2**), and (ii) statistical corroboration of the superadditive interaction. To assess the visual patterns in the results, **Figure 3** reports the means and estimated standard errors (± 1) for each condition, arranged in an interaction plot. We present the results two ways. The first is as z -score transformed scores (by participant), which reduces the impact of common forms of scale bias among the participants. The second is as raw results from the 7-point scale. Though we believe that the z -score transformed scores are likely the best option for analyzing acceptability judgments, as one anonymous reviewer points out, the raw scores allow us to evaluate the effect of the z -score transformation. The one caveat is that we must be sure not to exercise researcher degrees of freedom by selecting the results that we prefer. For this project, there is no risk – the z -score transformed scores and raw scores yield the same results.

For statistical corroboration, we conducted two types of analyses: one in a null hypothesis testing framework and one in a Bayesian framework. For the null hypothesis test, we constructed linear mixed effects models with SCRAMBLING and STRUCTURE as fixed effects and participant and item as random effects (intercepts only) for each island type using lme4 package in R (Bates et al. 2015). We calculated p -values using the lmerTest package, which uses the Satterthwaite approximation for degrees of freedom (Kuznetsova et al. 2017).⁷ We will interpret p -values below the conventional threshold of .05 as evidence against the null hypothesis, and therefore by implication, corroboration of the presence of an island effect. We will interpret p -values above the conventional threshold of .05 as a failure to reject the null hypothesis. Because the failure to reject the null hypothesis cannot be interpreted as evidence in support of the null hypothesis (because the null hypothesis is assumed to be true in the calculation of p -values), we include a Bayesian analysis to directly evaluate the null hypothesis.

For this Bayesian analysis, we calculated Bayes factors for the interaction term for the fixed effects in the linear models using the BayesFactor package (Morey & Rouder 2018). The Bayes factors reported here are of the BF_{10} type: they report the ratio of the likelihood of the data under the experimental hypothesis (H1) that an interaction is present to the likelihood of the data under the null hypothesis (H0) that there is no interaction present. Following Jeffreys (1939/1961), we will interpret a BF_{10} greater than 3 as strong evidence that an interaction is present, as this indicates that the data is at least 3-times more likely under a theory in which the interaction is present than one in which the interaction is absent. Similarly, we will interpret a BF_{10} less than 0.33 as strong evidence that there is no interaction, as this indicates that the data is 3-times

⁷ The full set of results are reported in the OSF repository for the current study (DOI [10.17605/osf.io/w2zdvt](https://doi.org/10.17605/osf.io/w2zdvt)).

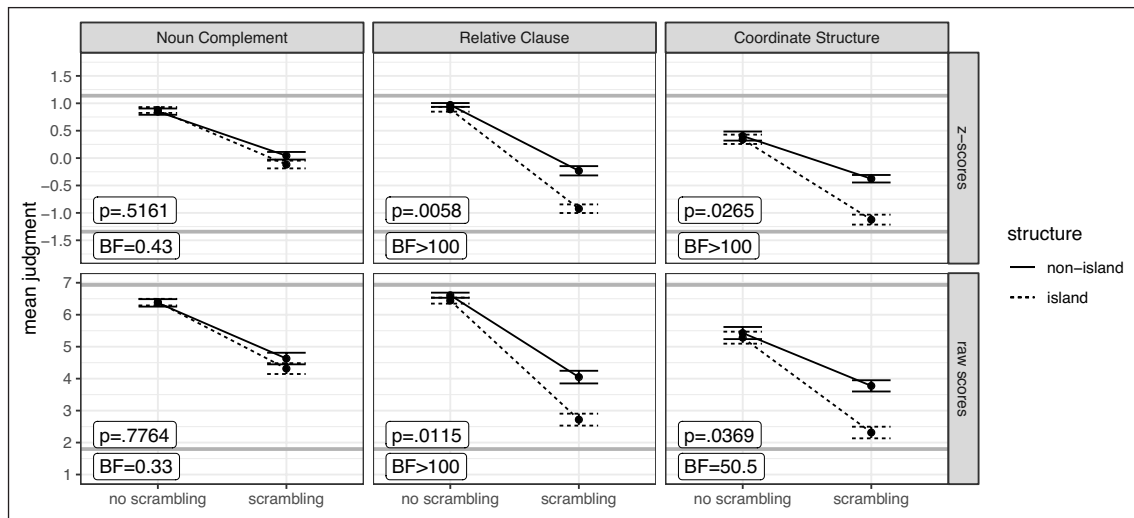


Figure 3: Experiment 1. Interaction plots for NP-scrambling in Japanese. Points are condition means. Error bars represent 1 estimated standard error in either direction. The top row reports the z-score transformed results, and the bottom row reports the raw results. The columns report each island type. The horizontal gray lines indicate the mean rating of the highest and lowest rated filler type to help assess ceiling and floor effects. For space reasons, p -values are rounded to a floor of .0001 and Bayes factors are rounded to a ceiling of 100.

more likely under the null hypothesis that the interaction is absent. We will also interpret Bayes factors between 0.33 and 3 as inconclusive (as the data is equally likely under both theories). In **Figure 3**, we have added the interaction term p -value and interaction BF_{10} to each cell of the plot so that the visual patterns and statistical results can be evaluated simultaneously.

As **Figure 3** makes clear, in Experiment 1, we see clear evidence of island effects with both relative clauses and coordinate structures – the visual pattern suggests a superadditive interaction, and both statistical analyses corroborate the interaction. However, for noun complements, we see no visual pattern of an interaction. The p -value is substantially above the conventional threshold, suggesting a failure to reject the null hypothesis. The Bayes factor is 0.43, which is close to the conventional threshold of 0.33, and suggests that the data is about 2.5-times more likely under the null hypothesis than under the experimental hypothesis. (The Bayes factor for the raw judgments is at 0.33, but we believe z-scores are the more principled analysis, and therefore focus on the statistical results for z-scores to avoid the appearance of leveraging researcher degrees of freedom to our benefit.) We also note that the mean rating of the island violating condition for noun complements is above the middle of the raw scale (above 4) and right at the middle of the z-score scale, which represents the mean judgment of all the items in the experiment (target conditions and fillers). This is noticeably different from the mean ratings for the island violating conditions for relative clauses and coordinate structures, as both are substantially below the midpoint of both the raw and z-score scales. We thus conclude that

Experiment 1 suggests that noun complements are not islands for NP-scrambling, while relative clauses and coordinate structures are. (We also note that there is no evidence of floor or ceiling effects as the mean ratings of all conditions lie below the means of the highest rated fillers and above the means of the lowest rated fillers.)

3.4.2 Variability in island effects between participants

As briefly discussed in Section 2.2, one possibility raised by Yano (2019) to explain the complex pattern of results for noun complements across the two experiments that he reports is that the island status of noun complements may show more between-participant variability than other island types. To investigate this, we calculated the size of the island effect reported by each participant as a differences-in-differences (DD) score (Maxwell & Delaney 2003): $(\text{non-island/scrambling} - \text{island/scrambling}) - (\text{non-island/no-scrambling} - \text{island/no-scrambling})$. These DD scores will be positive when the participant shows a superadditive interaction indicative of an island effect, with the magnitude indicating the size of the effect; these DD scores will be 0 when the participant shows no interaction, and negative if the participant shows a pattern in which the island-violating condition is more acceptable than the main effects of structure and scrambling would predict (this latter case is not predicted by any theory, so may be indicative of noise in that participant's responses).

Figure 4 reports the distribution of island effect sizes by participant as measured using DD scores for both z-scores and raw scores using histograms overlaid with probability density estimates.

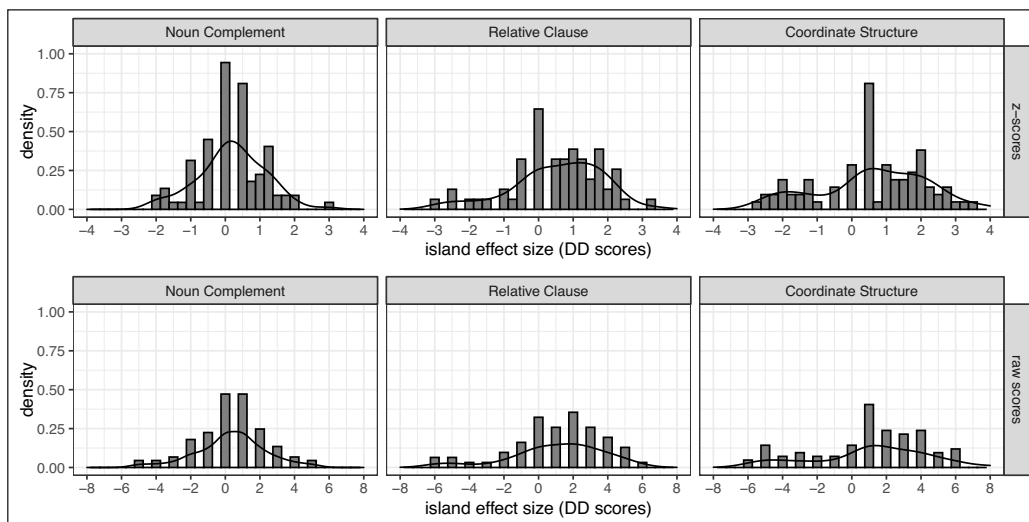


Figure 4: Experiment 1. The distribution of island effect sizes by participant, calculated as differences-in-differences scores for both z-scores and raw scores. The solid line is an estimate of probability density.

One clear sign that noun complements are more variable than the other islands would be for the distribution for noun complements to be wider than the distributions for the other islands. However, this is not what we see in **Figure 4**. If anything, noun complements show a narrower distribution. What we see instead is that noun complements show a relatively normal distribution centered exceedingly close to 0, as expected if there is no island effect, while the other two islands show distributions that are substantially shifted toward the positive range, as expected if there is an island effect. We therefore conclude that it is unlikely that noun complements show more between-participant variability than the other island types. Instead, we see further corroboration from the relatively normal, and relatively narrow, distribution for noun complements that there is no island effect.

4 Experiment 2

4.1 The logic of the design

For Experiment 2, we modified the design in two ways. First, we increased the number of tokens that participants rated per condition from one to two. This allows us to investigate not only the presence of island effects and the variability between participants as in Experiment 1, but also the variability of each condition within participants across the two ratings. We will therefore report three subsections of results for Experiment 2. Second, we used the same non-island structure for all three island types – specifically, an embedded declarative CP. This is a logically possible non-island structure for all three islands, therefore testing it adds a dimension of generality to our results. (The logic of the factorial design is such that the measurement of the island effect, which is in the interaction term, will not be affected by the choice of the non-island structure, as long as the non-island structure does not itself induce an interaction with scrambling. The only consequence of this change is in the main effect of structure.) Using the same non-island structure across all three islands also reduced the number of conditions tested (by 4), helping to offset the increase in tokens per condition.

4.2 Materials and survey construction

In Experiment 2, each participant completed a survey that consisted of 60 items: 16 experimental items (2 each of 8 target conditions) and 44 filler items pseudorandomized to avoid related experimental items appearing in succession. The 8 target conditions were non-scrambling and scrambling versions of declarative CPs (as in 8a and 8b), noun complements (8c and 8d), relative clauses (as in 9c and 9d), and coordinate structures (10c and 10d). We created 4 lexically matched sets of items per structure. The items were then distributed among 2 experimental lists using a Latin square procedure so that participants saw a unique lexical item in each trial.⁸

⁸ The complete lists of the experimental and filler sentences used in Experiment 2 can be found in the OSF repository for the current study (DOI [10.17605/osf.io/w2zdv](https://doi.org/10.17605/osf.io/w2zdv)).

4.3 Participants and presentation

Ninety-three participants from two universities in Tokyo, Japan, participated in the experiment. We excluded three participants from analysis because their answers to our language background questionnaire revealed that they had significant exposure to a language other than Japanese before they were 10 years old. Ninety self-reported native speakers of Japanese remain in the analysis. Participants either received course credit for their participation or 500 yen. The presentation was identical to Experiment 1.

4.4 Results

In this section we describe the results of the experiments, with a focus on the three questions licensed by the new design: (i) the presence or absence of the superadditive interaction indicative of island effects (and, relatedly, the relative size of the effect), (ii) the variability of island effects across participants, and (iii) the consistency of participants' ratings across the two tokens of each condition.

4.4.1 The presence or absence of island effects

As **Figure 5** shows, Experiment 2 again revealed clear evidence of island effects with coordinate structures and relative clauses – the visual pattern suggests a large superadditive interaction, and both statistical analyses corroborate the interaction. However, for our critical case, noun

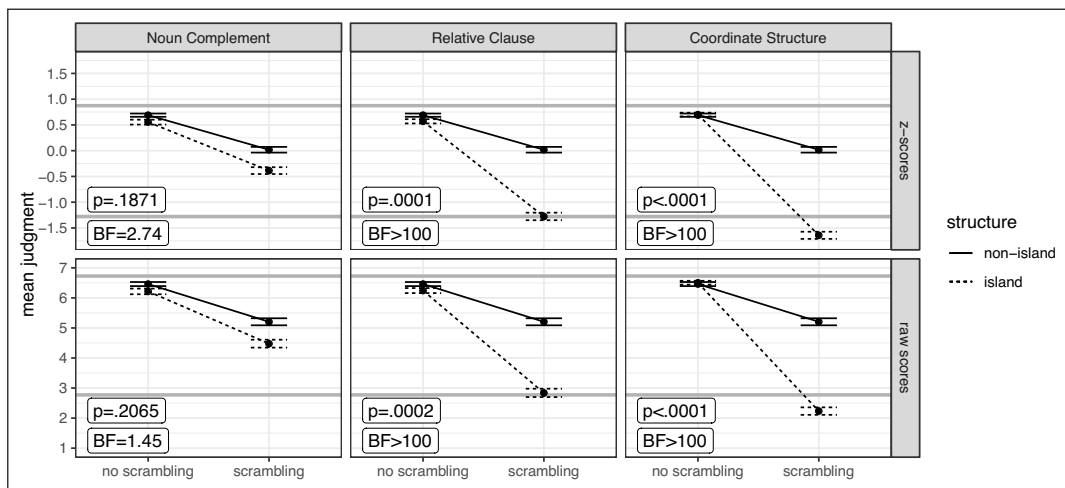


Figure 5: Experiment 2. Interaction plots for NP-scrambling in Japanese. Points are condition means. Error bars represent 1 estimated standard error in either direction. The top row reports the z-score transformed results, and the bottom row reports the raw results. The columns report each island type. The horizontal gray lines indicate the mean rating of the highest and lowest rated filler type to help assess ceiling and floor effects. For space reasons, p -values are rounded to a floor of .0001 and Bayes factors are rounded to a ceiling of 100.

complements, the pattern is more complicated. There is a small visual trend toward an effect, albeit with the potential island violating condition (island/scrambling) above the mid-point of the 7-point scale, and the two statistical tests trend in opposite directions: the p -value is well above the conventional threshold of .05, suggesting no evidence of an island effect, while the Bayes factor is trending toward the conventional threshold of 3 for the z -transformed results (although it is still technically below it).⁹

We prefer to be cautious in interpreting this pattern of results. The logically weakest conclusion we can make is that there is no clear evidence of an effect for noun complements, but we also cannot entirely rule it out. A hypothetical future experiment with even more observations per condition could potentially detect a very small effect using this design. We can also conclude that, if noun complements do have this hypothetical effect, it differs from the effects that we see for relative clauses and coordinate structures in important ways. For one, this hypothetical effect would be substantially smaller in size. It must be so small that it did not appear at all in Experiment 1 with 89 participants and one token per participant and does not reliably appear in Experiment 2 with 90 participants and two tokens per participant. In contrast, relative clauses and coordinate structures show relatively large and reliable effects in both experiments. (According to Sprouse & Almeida 2017 only 37 participants and one token per participant are necessary for 80% power to detect medium effect sizes and only 17 participants to detect large effect sizes.) For another, as the raw scores show, the island violating condition (island/scrambling) for noun complement islands is rated above the midpoint of the scale (around 4.5 in both experiments) and does not appear to result in unacceptability, in contrast with relative clause and coordinate structure islands (the island violating condition is in the lower half of the scale). We also note that the island/scrambling condition of the coordinate structure island is rated below the mean of the lowest rated filler. This means that it could be sitting at the floor in a world in which our lowest rated filler failed to be as low as the floor. But since coordinate structures show the largest island effect in the experiment, the theoretical consequence of underestimating its effect size is minimal. Instead, it tells us that the island/scrambling condition for relative clause islands is not sitting at the floor, despite being roughly equal in acceptability to the lowest rated filler. This underscores how much larger these island effects are compared to the hypothetical effect for noun complements, and how much lower the ratings of the island violating conditions are.

To summarize, by conventional statistical criteria, Experiment 2 provides strong evidence for large, classic island effects with relative clauses and coordinate structures, but no evidence either for or against island effects with noun complements. If one wishes to interpret the visual and BF trend as evidence that there may be a small effect of noun complements that we failed to detect,

⁹ The full set of results are reported in the OSF repository for the current study (DOI [10.17605/osf.io/w2zdvr](https://doi.org/10.17605/osf.io/w2zdvr)).

one must also conclude that it differs substantially from relative clause and coordinate structure islands both in size and in location in the scale.

4.4.2 Variability in island effects between participants

Turning next to the variability in island effects between participants in Experiment 2, **Figure 6** shows that noun complements once again show a relatively narrow normal distribution. There is no evidence that there is excessive variability in noun complements compared to relative clauses and coordinate structures. We do note, however, that the center of the distribution for noun complements is shifted slightly toward the positive, as expected given the small trend that we saw in the mean ratings in **Figure 5**. The two island types continue to show the same substantial shift toward the positive that we saw in Experiment 1.

4.4.3 Variability in island effects within participants

Though there is no evidence of increased variability for noun complements between participants, it is possible that there is increased variability within participants. Recent work by Kush et al. (2018; 2019) in Norwegian has suggested that some island effects that appear relatively small when viewed through the grand means of the sample may in fact be driven by inconsistent judgments within each participant.¹⁰ Though the source of this inconsistency is still an open area of investigation, here we provide a similar analysis for the ratings in Experiment 2.

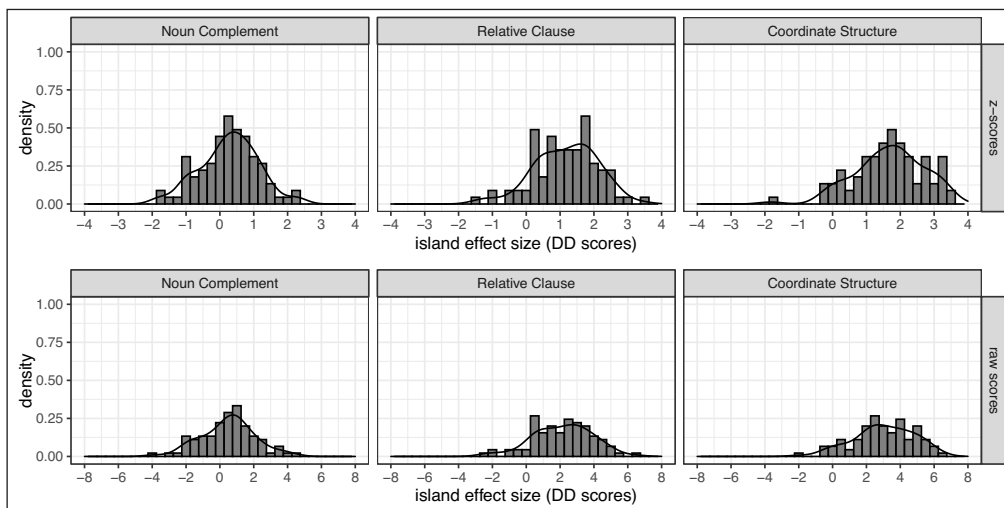


Figure 6: Experiment 2. The distribution of island effect sizes by participant, calculated as differences-in-differences scores for both z-scores and raw scores. The solid line is an estimate of probability density.

¹⁰ Kush & Dahl (2020) also found a similar case of inconsistent judgments in Norwegian-speaking English learners' judgments of *wh*-islands in English. We would like to thank an anonymous reviewer for directing our attention to this study.

Figure 7 plots the two judgments that each participant gave for each structure in a scatterplot, with the first judgment along the x-axis and the second judgment along the y-axis. The columns represent each structure, and the rows separate the no-scrambling (top row) and scrambling (bottom row) conditions. We divide each plot into four quadrants. A point in the top right quadrant (Quadrant 1) represents a participant who rated both tokens in the upper half of the scale. For convenience, we will label such a pattern *consistent acceptor*. A point in the bottom left quadrant (Quadrant 3) represents a participant who rated both tokens in the lower half of the scale. We can label such a pattern *consistent rejector*. The other two quadrants (Quadrants 2 and 4) represent participants who rated one token in the upper half of the scale and one token the lower half of the scale. We will label this pattern *inconsistent*. We have added two features to make the plot a bit easier to read: colors representing the three patterns, and two-dimensional (joint) probability density estimates to draw attention to the density of the points in each location. Similar to a topographic map, in a two-dimensional probability density plot, concentric circles that are closer together represent higher density (because, like topographic maps, these plots are looking down on the peaks in the density space from directly above).

Figure 8 presents the same plot for the raw scores. We include the raw scores here because we saw in **Figure 5** that the midpoint of the z-score scale corresponded with ratings above the midpoint of the raw score scale. This suggests that, on average, the balance of items in the survey was slightly skewed toward higher ratings. This means that the dividing lines in **Figure 7** represent consistency relative to the midpoint of the distribution of items in the survey, while

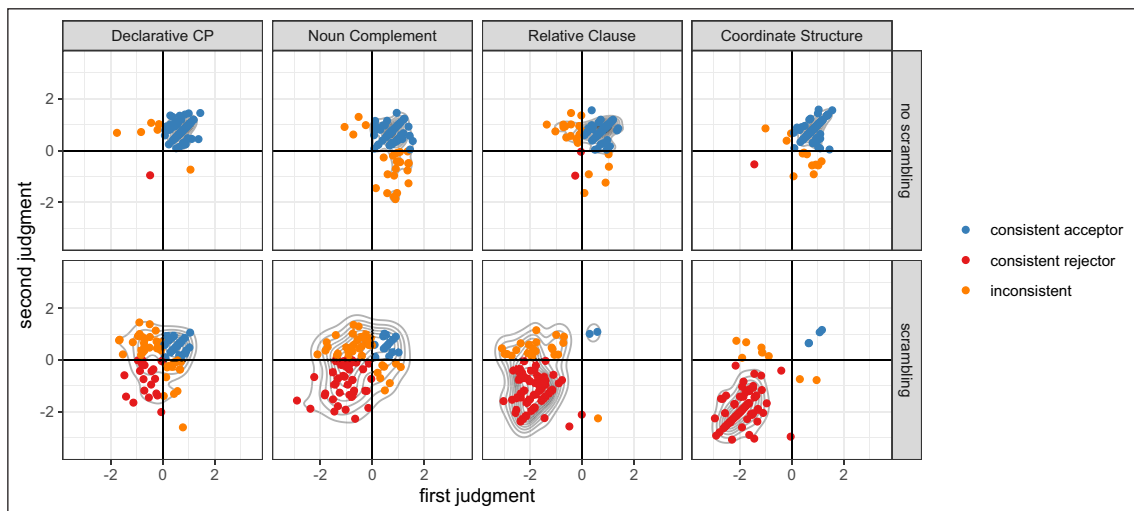


Figure 7: Experiment 2. Scatterplots of the ratings for the two tokens of each condition for each participant, with two-dimensional (joint) probability density estimates overlaid. The points are colored according to the type of judgment pattern defined by the midpoint (0) of the z-score scales.

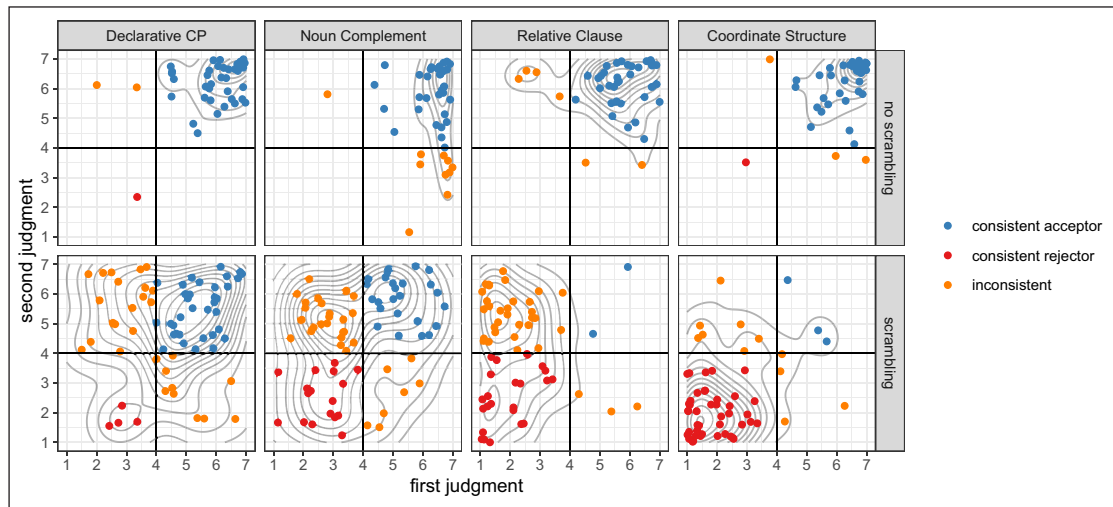


Figure 8: Experiment 2. Scatterplots of the ratings for the two tokens of each condition for each participant, with two-dimensional (joint) probability density estimates overlaid. The points are colored according to the type of judgment pattern defined by the midpoint (4) of the raw score scales. The integer nature of the raw scale would normally mean that many points perfectly overlap. We have added a small amount of jitter to the points to make them all visible.

the dividing lines for the raw scores in **Figure 8** represent consistency relative to the absolute midpoint of the raw rating scale. We have added a small amount of jitter to the points in **Figure 8** to make all of the points visible (without jitter, there is quite a bit of overlap among points because the raw scale only allows 7 distinct ratings).

In both figures we see the same patterns. The no-scrambling conditions (top row) appear to show the upper bounds of consistency in this experiment – the vast majority of participants show the consistent acceptor pattern (Quadrant 1) for each structure, with a small number of inconsistent patterns mixed in. In contrast, the scrambling conditions (bottom row) reveal potentially relevant patterns. In the first column, for the by-hypothesis grammatical scrambling out of declarative CP, we see that the largest mass of participants is in the consistent acceptor quadrant, while there is also a non-negligible number of participants in each of the other three quadrants. This provides more nuance to the mean rating in **Figure 5** – we see now that the middle-of-the-scale mean rating (near 0) was actually driven by a mix of consistent acceptors, consistent rejectors, and inconsistent participants. This sets a baseline expectation for NP-scrambling consistency: the rating of NP-scrambling itself, in the absence of islands, is relatively variable in Japanese. We can then apply this baseline as we look at the potential island structures.

For noun complements, we see a shift in the probability mass that moves a bit to the left and down from the declarative CP baseline. The center of mass does not quite fully cross into the consistent rejector quadrant, but instead hovers over the horizontal axis line (indicating a rating near 0 for the second token). This shift from the baseline established by scrambling out

of declarative CPs is in line with the equivocal results that we saw for noun complements in the means in **Figure 5** and the DD scores in **Figure 6** – there is a small trend toward a slightly negative rating, but enough variability that it is still plausibly an effect around 0. For relative clauses, we see a shift further left and down. Though there is still variability in relative clauses, the vast majority of participants are either consistent rejectors or inconsistent raters, consistent with the grand means in **Figure 5** and DD scores in **Figure 6**. Finally, we see an even further shift toward the consistent rejector quadrant for coordinate structures, which is again consistent with the grand means in **Figure 5** and the DD scores in **Figure 6**.

Taken together, what we see is that scrambling itself introduces a fair amount of within-participant variability to judgments when compared to no-scrambling conditions. Noun complements appear (visually) to show roughly the same amount of variability as scrambling from declarative CP structures. This leads to two conclusions. The first is that there is no additional variability in noun complements that could explain past debates about their island status. The second is that noun complements show the same general pattern of variability as unequivocally grammatical sentences. This stands in contrast to the unequivocal islands, which tend to show a mild increase in consistency compared to the grammatical controls. This again points to noun complements being qualitatively distinct from the other islands (and perhaps more similar to grammatical sentences), and thus further corroborates our conclusion that noun complements in Japanese are not islands for NP-scrambling.

5 General Discussion and Implications

This paper presented two experiments using the factorial definition of island effects to compare three island types with relatively large samples (89 and 90 participants respectively): relative clauses, noun complements, and coordinate structures. Experiment 1 tested one token per condition, whereas Experiment 2 tested two tokens per condition, allowing for an investigation of within-participant variability. Both experiments unequivocally show that relative clauses and coordinate structures are islands for NP-scrambling in Japanese, corroborating previous studies (e.g., Harada 1977; Haig 1976; Saito 1985; 1987). Experiment 1 showed a clear lack of evidence of noun complement island effects, while the results from Experiment 2 were statistically inconclusive and fail to rule out the possibility that there is a small effect for noun complements. However, these effects are substantially smaller than relative clause and coordinate structure islands, and they are also qualitatively different as the potential island violating condition with noun complements is above the midpoint of the scale, unlike the island violating condition for relative clauses and coordinate structures. We also closely examined our results for between- and within-participant variability. Although Yano (2019) raised the possibility that noun complements are associated with an increase in between-participant variability (compared to adjunct islands), we found that noun complements show the same or less between-participant variability than

relative clauses and coordinate structures. We also found that noun complements show the same amount of within-participant variability as unequivocally grammatical scrambling conditions. Taken as a whole, we interpret our findings to suggest that noun complements are not islands (joining Haig 1976). However, the hypothetical small effect observed with noun complements in Experiment 2 potentially explains why some researchers have reported mild noun complement island effects in the past (e.g., Saito 1985; 1987), or observed inconsistent results with typical sample sizes (e.g., Yano 2019).

Though it has occasionally been claimed that island effects for noun complements are smaller than island effects with relative clauses (e.g., Chomsky 1986), to the best of our knowledge, our study is the first to provide formal experimental evidence that the two islands pattern qualitatively differently. In fact, the only previous studies to our knowledge to directly test both noun complements and relative clauses in the same formal experiment are Kush et al. (2018; 2019), which tested both islands in Norwegian for *wh*-questions and topicalization, respectively. The results of these two studies suggest that the effect size for noun complements and relative clauses are approximately equal in Norwegian (though it is always possible that the difference in effect size is simply too small to detect reliably with the sample sizes used in these studies). Our findings also challenge the claim that noun complements are simply relative clauses in disguise (e.g., Nichols 2003; Kayne 2008; 2010; Arsenijević 2009; Haegeman 2012; cf. de Cuba 2017), and the claim that Japanese lacks English-like relative clauses entirely (e.g., Kuno 1973; Murasugi 2000). Under these analyses, it would be unexpected to find that only relative clauses are islands in Japanese.

We offer two more observations about the two types of complex NPs in Japanese. First, our results appear to challenge Stowell's (1981) suggestion that island status correlates with the possibility of complementizer deletion: CP complements of verbs allow complementizer deletion (11) and are not islands, while CP complements of nouns do not allow complementizer deletion and are islands (12). (The examples in (11) and (12) are our own, with diacritics indicating the pattern discussed by Stowell 1981.)

(11) Jessica claimed that/ \emptyset Lisa invented the app.

(12) Jessica made the claim that/* \emptyset Lisa invented the app.

Stowell argues that complementizer deletion is possible in (11) because the embedded clause is a true complement of the verb (and therefore the empty category created by the deletion is governed, satisfying the ECP). Stowell further argues that the impossibility of complementizer deletion in (12) suggests that the embedded clause is, in fact, not a complement, but rather an adjunct (leading to an ECP violation because the empty category created by the deletion is not governed; cf. Grimshaw 1990; Kiss 1990; Takahashi 1994; Sabel 2002; de Cuba 2017).

If the embedded clause in noun complement constructions is in fact an adjunct, the island effect observed with English noun complements is expected as a type of adjunct island effect. Interestingly, Fukui (1988) argues that the Japanese noun complements with *toyuu* also involve an adjunct because *toyuu* cannot be deleted, as in (13).

- (13) Taro-ga sore-o teniire-ta toyuu/* \emptyset uwasa
 T-NOM it-ACC obtain-PST that.say rumor
 ‘the rumor that Taro obtained it’ (Fukui 1998: 513; (26))

If Fukui (1988) is correct, and the embedded clause inside the noun complement is an adjunct, our finding that *toyuu* noun complements show little to no island effect is puzzling. Fukui’s claim, however, is not supported by other Japanese-internal facts about complementizer deletion. First, the Tokyo dialect does not allow for deletion of the complementizer *to/tte* even when the embedded clause appears to be a clear case of a verbal complement. (The example in (14) is our own, with diacritic indicating the pattern reported in Fukuda 2000).

- (14) Taro-wa Hanako-ga ki-ta to/tte/* \emptyset it-ta/omot-ta.
 T-TOP H-NOM come-PST TO/TTE say-PST/think-PST
 ‘Taro thought/said (that) Hanako came.’

Thus, we have no reason to expect that complementizers in general can be deleted even when the embedded clause is indeed a complement; therefore, it is not surprising that *toyuu* in (13) cannot be deleted. Second, in some Western dialects of Japanese, the complementizer deletion in (14) is acceptable (Saito 1987; Fukuda 2000; Kishimoto 2006). Yet, crucially, no one has claimed that the CP complement of a verb with *to* in the Tokyo dialect is an island while it is a non-island in these Western dialects. What is more, according to speakers of these dialects whom we consulted, the same complementizer *to/tte* can also be deleted in noun complements, as in (15), though we note that this fact should be quantified in future judgment or corpus studies.

- (15) Taro-ga sore-o teniire-ta tte/ \emptyset yuu uwasa
 T-NOM that-ACC obtain-PST TTE YUU rumor
 ‘the rumor that Taro obtained that’

Thus, when we look at the facts of complementizer deletion across both English and Japanese, the apparent correlation between island status and the possibility of complementizer deletion disappears.¹¹ An anonymous reviewer also notes that Spanish provides further evidence against

¹¹ As noted in Section 2.1, we are following Fukui’s assumption that *toyuu* is an English-style complementizer similar to *that*. It could be argued that *toyuu* is not a complementizer, but rather a semi-frozen verbal element similar to the English noun phrase “the rumor that *says that* Taro obtained it.” In that case, the example in (13) may no longer be relevant to the correlation between complementizer deletion and island status. But, to our knowledge, it is widely agreed that *to* is a complementizer in Japanese; therefore, the examples in (14) and (15) are still relevant to the correlation.

half of the correlation with complementizer deletion. Pañeda et al. (2020) show that there is either no island effect or a very small island effect for noun complements in Spanish. Noun complements in Spanish do not allow complementizer deletion, so under the Stowell (1981) analysis, they should show island effects, contrary to fact.

Our second observation is that island status does not appear to correlate with the syntactic complexity of the embedded clause (in terms of number of available positions in the clause, or amount of functional structure within the clause). While relative clauses and noun complements in English are analyzed as involving an embedded CP, i.e., embedded clauses with the same complexity, as discussed in Section 2.1, there are several reasons to believe that the structure of relative clauses in Japanese is less complex along certain dimensions than the structure of the embedded clauses inside noun complements. Despite this observation, only relative clauses show clear island effects. This observation potentially has implications for bounding-based approaches to island effects (like the classic Subjacency and barriers approaches). Though it clearly depends on the details of the theory, in principle, more complex syntactic constituents like noun complements in Japanese have the potential to host more bounding nodes or barriers than less complex constituents like relative clauses, despite island effects patterning in the opposite direction. This suggests that our results may be relevant for adjudicating among specific implementations of bounding-based approaches to island effects.

While it is beyond the scope of this paper to evaluate the full set of theories of islands in the literature, we would like to mention the consequences of our results for a few prominent theories to illustrate their potential theoretical value. Huang's (1982) Condition on Extraction Domains, Lasnik & Saito's (1984) gamma-marking, and Chomsky's (1986) barriers approach all share the intuition that there is a fundamental distinction between adjunct CPs and complement CPs (in terms of government, gamma-marking, and L-marking respectively), and that this distinction causes (most) adjuncts to be islands and (most) complements to be non-islands. English noun complements create complications for this view, as they appear to be complements, but are nonetheless islands. Under these approaches, our finding that Japanese noun complements are likely not islands suggests that they are indeed complements, and that it is English noun complements that are exceptional in some way, not noun complement constructions in general (but see Hankamer & Mikkelsen 2021 for a proposal that Dutch and English noun complements involve a CP complement to D or a CP adjunct to DP). The intuition that complements and adjuncts are fundamentally distinct, and that this difference is the source of island effects, is also central to modern phase-based approaches to island effects. For example, Rackowski & Richards (2005) propose that phrases that enter into an Agree relation with a phase head (i.e., v) are transparent to extraction, while phrases that do not enter into an Agree relation are islands. They present evidence that, in Tagalog, complement CPs, which are transparent to extraction, show morphological evidence of this Agree relation. Under this approach, our results suggest

that, while noun complements in English must not enter into an Agree relation with the next phase head, noun complements in Japanese do, despite the fact that neither language shows a morphological reflex of this relation. Similarly, Müller (2010) proposes that phases are transparent to extraction as long as they are active and can thus be given an edge feature to accommodate the extraction, in turn circumventing the Phase Impenetrability Condition of Chomsky (2001; 2008). Phase heads are active until their final (i.e., last merged) specifier is merged. Adjuncts are islands because they are the final specifier of a phase head, while complements are not islands because (by definition) they are not specifiers. Under this approach, English noun complements must be final specifiers, and thus adjuncts, while Japanese noun complements would be true complements. For each of these potential analyses, future studies could explore the properties of English and Japanese noun complements (beyond complementizer deletion) to determine if there is independent evidence (beyond island effects) for postulating these critical differences between the two types of CPs.

6. Conclusion

This article presented two studies that contribute to the discussion of complex NP islands in Japanese. While there is little contention among previous studies that relative clauses are islands for NP-scrambling, there is no consensus regarding noun complements: Haig (1976) argued they are not islands, Saito (1985; 1987) claimed they are, and Yano's (2019) two experiments yielded potentially conflicting results, with one showing a small island effect and the other showing no island effect. We presented two acceptability judgment experiments to compare relative clauses and noun complements with coordinate structures, which are uncontroversially islands. Both experiments were designed using the factorial definition of island effects. Participants saw one token per condition per participant in Experiment 1, while participants saw two tokens per condition in Experiment 2, thus allowing us to explore possible between- and within-participant variability, which has been raised as a possible complicating factor in previous studies of island effects (Kush et al. 2018; 2019; Yano 2019). Our results corroborated previous studies in that relative clauses and coordinate structures yield large island effects in Japanese. Our results for noun complements yielded conclusive evidence that they are not islands in Experiment 1, and yielded the weaker conclusion that noun complements are quantitatively and qualitatively distinct from relative clauses and coordinate structures in Experiment 2. A closer look at the results revealed no between-participant variability in the judgments on noun complements in Experiments 1 or 2, and the same level of within-participant variability (in Experiment 2) for both scrambling out of declarative CPs and noun complements. Therefore, we conclude that noun complements in Japanese are not islands the way that relative clauses are, but leave open the possibility that noun complements in Japanese yield extremely small effects that cannot be reliably detected even with the large sample sizes here.

Our results suggest a clear difference between relative clauses and noun complements in Japanese. Furthermore, our results suggest a clear difference between noun complements in Japanese and noun complements in other languages that have been tested using the factorial definition, such as English, Italian, and Norwegian, which show large, robust island effects (Sprouse et al. 2011, Sprouse et al. 2012, Sprouse et al. 2016, Kush et al. 2018; 2019). We have argued that these findings have direct consequences for most existing theories of island effects, for theories of the relationship between complementizer deletion and island status, and for theories of the relationship between the complexity of syntactic structure and island status. This in turn suggests that future studies should probe the properties of relative clauses and noun complements (cross-linguistically) along these dimensions.

Our study used scrambling dependencies to evaluate island effects, because *wh*-questions, which are typically used to investigate island effects in *wh*-movement languages, do not move overtly in Japanese. There are, however, studies that suggest that *wh-in-situ* also plays an important role in the theories of islands in languages without overt *wh*-movement. Studies such as Tanaka & Schwartz (2018) on Japanese and Lu et al. (2020) on Mandarin Chinese found that argument *wh*-phrases that stay *in situ* within a relative clause give rise to island effects, contrary to previous claims that only certain *wh*-adjuncts invoke island effects in these languages. (cf. Kim & Goodall 2016, whose experimental investigation of island effects involving *wh-in-situ* in Korean also observed island effects; but cf. Sprouse et al. 2011, who found no island effects for subject, adjunct, *whether*, and noun complement islands in Japanese.) The complex picture emerging from these studies suggests a need for comprehensive comparisons across island effects and dependency types within *wh-in-situ* languages.

Abbreviations

ACC = accusative, CNP = complex noun phrase, COMP = complementizer, CP = complementizer phrase, DAT = dative, GEN = genitive, NC = noun complement, NOM = nominative, NPST = non-past, PST = past, RC = relative clause, TOP = topic

Additional files

The additional files for this article can be found as follows:

- **Appendices:** Lists of the materials for Experiments 1 and 2 and the full sets of the results of statistical analysis (doi [10.17605/osf.io/w2zdvr](https://doi.org/10.17605/osf.io/w2zdvr))

Acknowledgements

Parts of this study were presented at the 27th Japanese/Korean Linguistics Conference at Sogang University and at the Tuesday Seminar at University of Hawai'i at Mānoa. We would like to thank the audiences at these events for their helpful feedback and comments. We would also like to thank Kamil Ud Deen and Li “Julie” Jiang for reading earlier drafts of this study and three anonymous reviewers for *Glossa* and the handling editor Lyn Tieu for their feedback and comments that improved various aspects of this study. Finally, we would like to thank Yuki Hirose for her generous help in recruiting participants for our experiments. All remaining errors are our own.

Competing interests

The authors have no competing interests to declare.

References

- Almeida, Diogo. 2014. Subliminal *wh*-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista da ABRALIN* 13(2). 55–93. DOI: <https://doi.org/10.5380/rabl.v13i2.39611>
- Arsenijević, Boban. 2009. Clausal complementation as relativization. *Lingua* 119(1). 39–50. DOI: <https://doi.org/10.1016/j.lingua.2008.08.003>
- Bates, Douglas & Mächler, Martin & 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>
- Chomsky, Noam. 1986. *Barriers*. Cambridge, Massachusetts: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In Kenstowicz, Michael (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, Massachusetts: MIT Press.
- Chomsky, Noam. 2008. On phases. In Freidin, Robert, Otero, Carlos P. & Zubizarreta, Maria Luisa (eds.), *Foundational issues in linguistic theory*, 133–166. Cambridge, Massachusetts: MIT Press.
- Christensen, Ken Ramshøj & Kizach, Johannes & Nyvad, Anne Mette. 2013. Escape from the island: Grammaticality and (reduced) acceptability of *wh*-island violations in Danish. *Journal of Psycholinguistic Research* 42. 51–70. DOI: <https://doi.org/10.1007/s10936-012-9225-3>

- de Cuba, Carlos. 2017. Noun complement clauses as referential modifiers. *Glossa: a journal of general linguistics* 2(1). 3. DOI: <https://doi.org/10.5334/gjgl.53>
- Drummond, Alex. 2013. Ibex Farm. <https://spellout.net/ibexfarm/>.
- Dryer, Matthew S. 1980. The positional tendencies of sentential noun phrases in universal grammar. *The Canadian Journal of Linguistics* 25(2). 123–195. DOI: <https://doi.org/10.1017/S0008413100009373>
- Fukuda, Minoru. 2000. Complementizer drop and IP-complementation in Japanese. *Kansas Working Papers in Linguistics* 25. 39–52. DOI: <https://doi.org/10.17161/KWPL.1808.650>
- Fukui, Naoki. 1988. LF extraction of *naze*: Some theoretical implications. *Natural Language and Linguistic Theory* 6(4). 503–526. DOI: <https://doi.org/10.1007/BF00134490>
- Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, Massachusetts: MIT Press.
- Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena, and the composition of the left periphery*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199858774.001.0001>
- Haig, John H. 1976. Shadow pronoun deletion in Japanese. *Linguistic Inquiry* 7(2). 363–371.
- Hankamer, Jorge & Mikkelsen, Line. 2021. CP compliments to D. *Linguistic Inquiry* 52(3). 473–518. DOI: https://doi.org/10.1162/ling_a_00387
- Harada, Shin-Ichi. 1977. Nihongo ni “henkei” wa hitsuyô da [Japanese needs transformational rules]. *Gengo [Languages]* 6(11). 88–95 and 12. 96–103.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511554285>
- Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, Massachusetts: Massachusetts Institute of Technology dissertation.
- Jeffreys, Harold. 1939/1961. *Theory of probability*. Oxford: Oxford University Press.
- Josephs, Lewis S. 1976. Complementation. In Shibatani, Masayoshi (ed.), *Syntax and Semantics, vol. 5: Japanese Generative Grammar*, 307–369. New York: Academic Press. DOI: https://doi.org/10.1163/9789004368835_009
- Kayne, Richard. 2008. Antisymmetry and the lexicon. *Linguistic Variation Yearbook* 8(1). 1–32. DOI: <https://doi.org/10.1075/livy.8.01kay>
- Kayne, Richard. 2010. *Comparisons and contrasts*. Oxford: Oxford University Press.
- Keshev, Maayan & Meltzer-Asscher, Aya. 2018. A processing-based account of subliminal wh-island effects. *Natural Language & Linguistic Theory* 37(2). 621–657. DOI: <https://doi.org/10.1007/s11049-018-9416-1>
- 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>
- Kishimoto, Hideki. 2006. On the existence of null complementizers in syntax. *Linguistic Inquiry* 37(2). 339–345. DOI: <https://doi.org/10.1162/ling.2006.37.2.339>
- Kiss, É. Katalin. 1990. Why noun-complement clauses are barriers. In Mascaró, Joan & Nespó, Marina (eds.), *Grammar in progress: GLOW essays for Henk van Riemsdijk*, 265–278. Dordrecht: Foris. DOI: <https://doi.org/10.1515/9783110867848.265>

- Kluender, Robert & Kutas, Marta. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8(4). 573–633. DOI: <https://doi.org/10.1080/01690969308407588>
- Ko, Heejeong & Chung, Han-byul & Kim, Kitaek & Sprouse, Jon. 2019. An experimental study on scrambling out of islands: To the left and to the right. *Language and Information Society* 37. 287–323. DOI: <https://doi.org/10.29211/soli.2019.37..008>
- Koizumi, Masatoshi & Imamura, Satoshi. 2017. Interaction between syntactic structure and information structure in the processing of a head-final language. *Journal of Psycholinguistic Research* 46. 247–269. DOI: <https://doi.org/10.1007/s10936-016-9433-3>
- Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, Massachusetts: MIT Press.
- Kush, Dave & Dahl, Anne. 2020. L2 transfer of L1 island-sensitivity: The case of Norwegian. *Second Language Research*. 1–32. DOI: <https://doi.org/10.1177/0267658320956704>
- Kush, Dave & Lohndal, Terje & Sprouse, Jon. 2018. Investigating variation in island effects: A case study of Norwegian *wh*-extraction. *Natural Language and 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(3). 393–420. DOI: <https://doi.org/10.1353/lan.2019.0051>
- Kuznetsova, Alexandra & Brockhoff, Per B. & Christensen, Rune H. B. 2017. lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: <https://doi.org/10.18637/jss.v082.i13>
- Lasnik, Howard & Saito, Mamoru. 1984. On the nature of proper government. *Linguistic Inquiry* 15(2). 235–289.
- Lu, Jiayi & Thompson, Cynthia K. & Yoshida, Masaya. 2020. Chinese *wh-in-situ* and island: A formal judgment study. *Linguistic Inquiry* 51(3). 611–623. DOI: https://doi.org/10.1162/ling_a_00343
- Maxwell, Scott E. & Delaney, Harold D. 2003. *Designing experiments and analyzing data: A model comparison perspective*. London: Routledge. DOI: <https://doi.org/10.4324/9781410609243>
- Miyagawa, Shigeru. 2001. The EPP, scrambling, and *wh-in-situ*. In Kenstowicz, Michael (ed.), *Ken Hale: A life in language*, 293–338. Cambridge, Massachusetts: MIT Press.
- Morey, Richard D. & Rouder, Jeffrey N. 2018. BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12 <https://CRAN.Rproject.org/package=BayesFactor>
- Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41(1). 35–82. DOI: <https://doi.org/10.1162/ling.2010.41.1.35>
- Murasugi, Keiko. 2000. Japanese complex noun phrases and the Antisymmetry theory. In Martin, Roger & Michaels, David & Uriagereka, Juan & Keyser, Samuel Jay (eds.), *Step by step. Essays on minimalist syntax in honor of Howard Lasnik*, 211–234. Cambridge, Massachusetts: MIT Press.
- Nakau, Minoru. 1973. *Sentential complementation in Japanese*. Tokyo: Kaitakusha.
- Nemoto, Naoko. 1993. *Chain and case position: A study from scrambling in Japanese*. Storrs, Connecticut: University of Connecticut dissertation.

- Nemoto, Naoko. 1999. Scrambling. In Tsujimura, Natsuko (ed.), *The handbook of Japanese linguistics*, 121–153. Oxford: Blackwell. DOI: <https://doi.org/10.1002/9781405166225.ch5>
- Nichols, Lynn. 2003. Attitude evaluation in complex NPs. In Carnie, Andrew & Harley, Heidi & Willie, MaryAnn (eds.), *Formal approaches to function in grammar: In honor of Eloise Jelinek*, 155–164. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.62.12nic>
- Ogihara, Toshiyuki. 2004. Adjectival relatives. *Linguistics and Philosophy* 27. 557–608. DOI: <https://doi.org/10.1023/B:LING.0000033856.09799.c5>
- Omaki, Akira & Fukuda, Shin & Nakao, Chizuru & Polinsky, Maria. 2020. Subextraction in Japanese and subject-object symmetry. *Natural Language and Linguistic Theory* 38. 627–669. DOI: <https://doi.org/10.1007/s11049-019-09449-8>
- Pañeda, Claudia & Lago, Sol & Vares, Elena & Veríssimo, João & Felser, Claudia. 2020. Island effects in Spanish comprehension. *Glossa: A journal for general linguistics* 5(1):21. 1–30. DOI: <https://doi.org/10.5334/gjgl.1058>
- Rackowski, Andrea & Richards, Norvin. 2005. Phase edge and extraction: A Tagalog case study. *Linguistic Inquiry* 36(4). 565–599. DOI: <https://doi.org/10.1162/002438905774464368>
- Sabel, Joachim. 2002. A minimalist analysis of syntactic islands. *The Linguistic Review* 19(3). 271–315. DOI: <https://doi.org/10.1515/tlir.2002.002>
- Saito, Mamoru. 1985. *Some asymmetries in Japanese and their theoretical implications*. Cambridge, Massachusetts: Massachusetts Institute of Technology dissertation.
- Saito, Mamoru. 1987. Three notes on syntactic movement in Japanese. In Imai, Takashi & Saito, Mamoru (eds.), *Issues in Japanese Linguistics*, 301–350. Dordrecht: Foris. DOI: <https://doi.org/10.1515/9783112420423-011>
- Saito, Mamoru. 1989. Scrambling as semantically vacuous A'-movement. In Baltin, Mark R. & Kroch, Anthony S. (eds.), *Alternative conception of phrase structure*, 189–200. Chicago, Illinois: The University of Chicago Press.
- Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1. 69–118. DOI: <https://doi.org/10.1007/BF00129574>
- Sprouse, Jon. 2007. *A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge*. College Park, Maryland: University of Maryland, College Park dissertation.
- 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 & Caponigro, Ivano & Greco, Ciro & Cecchetto, Carlo. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34. 307–344. DOI: <https://doi.org/10.1007/s11049-015-9286-8>
- Sprouse, Jon & Fukuda, Shin & Ono, Hajime & Kluender, Robert. 2011. Reverse island effects and the backward search for a licenser in multiple *wh*-questions. *Syntax* 14(2). 179–203. DOI: <https://doi.org/10.1111/j.1467-9612.2011.00153.x>

- Sprouse, Jon & Wagers, Matthew & Phillips, Collin. 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 & Mušič, Manca & Stateva, Penka. 2018. Two (non-) islands in Slovenian: A study in experimental syntax. *Linguistics* 56(3). 435–476. DOI: <https://doi.org/10.1515/ling-2018-0002>
- Stowell, Tim. 1981. Complementizers and the Empty Category Principle. In Burke, Victoria & Pustejovsky, James (eds.), *Proceedings of NELS 11*, 345–363. Amherst, Massachusetts: University of Massachusetts, Graduate Linguistics Students Association.
- Szabolcsi, Anna & Lohndal, Terje. 2017. Strong vs. weak islands. In Everaert, Martin & van Riemsdijk, Henk C. (eds.), *Wiley Blackwell companion to syntax*, 2nd edn., 479–531, Oxford: Blackwell. DOI: <https://doi.org/10.1002/9780470996591.ch64>
- Tada, Hiroaki. 1993. *A/A-bar partition in derivation*. Cambridge, Massachusetts: Massachusetts Institute of Technology dissertation.
- Takahashi, Daiko. 1994. *Minimality of movement*. Storrs, Connecticut: University of Connecticut dissertation.
- Tanaka, Nozomi & Schwartz, Bonnie D. 2018. Investigating relative clause island effects in native and nonnative adult speakers of Japanese. In Bertolini, Anne B. & Kaplan, Maxwell J. (eds.), *Proceedings of the 42nd annual Boston University Conference on Language Development*, vol.2, 750–763. Somerville, Massachusetts: Cascadilla Press.
- Teramura, Hideo. 1982. *Nihongo no sintakusu to imi I* [Japanese syntax and semantics I]. Tokyo: Kuroshio.
- Tomioka, Satoshi. 2015. Embedded *wa*-phrases, predication and judgment theory. *Natural Language and Linguistic Theory* 33. 267–305. DOI: <https://doi.org/10.1007/s11049-014-9258-4>
- Tucker, Matthew A. & Idrissi, Ali & Sprouse, Jon & Almeida, Diogo. 2019. Resumption ameliorates different islands differentially: Acceptability data from Modern Standard Arabic. In Khalfaoui, Amel & Tucker, Matthew A. (eds.), *Perspectives on Arabic Linguistics XXX: Papers from the annual symposia on Arabic Linguistics, Stony Brook, New York, 2016 and Norman, Oklahoma, 2017*, 159–193. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/sal.7.09tuc>
- Yamashita, Hiroko. 2002. Scrambled sentences in Japanese: Linguistic properties and motivations for production. *Text* 22(4). 597–633. DOI: <https://doi.org/10.1515/text.2002.023>
- Yamashita, Hiroko & Chang, Franklin. 2001. Long-before-short preference in the production of a head final language. *Cognition* 81(2): B45–B55. DOI: [https://doi.org/10.1016/S0010-0277\(01\)00121-4](https://doi.org/10.1016/S0010-0277(01)00121-4)
- Yano, Masataka. 2019. On the nature of the discourse effect on extraction in Japanese. *Glossa: a journal of general linguistics* 4(1). 90. DOI: <https://doi.org/10.5334/gjgl.822>
- Yoshimura, Noriko. 1992. *Scrambling and anaphora in Japanese*. Los Angeles, California: University of Southern California dissertation.

