

## RESEARCH

## Island effects in Spanish comprehension

Claudia Pañeda<sup>1</sup>, Sol Lago<sup>2</sup>, Elena Vares<sup>3</sup>, João Veríssimo<sup>4</sup> and Claudia Felser<sup>4</sup><sup>1</sup> University of Oviedo, Oviedo, ES<sup>2</sup> Goethe University Frankfurt, Frankfurt am Main, DE<sup>3</sup> Escuelas Universitarias Gimbernat-Cantabria, Torrelavega, ES<sup>4</sup> Potsdam Research Institute for Multilingualism, University of Potsdam, Potsdam, DECorresponding author: Claudia Pañeda ([panedaclaudia@uniovi.es](mailto:panedaclaudia@uniovi.es))

A growing body of experimental syntactic research has revealed substantial variation in the magnitude of island effects, not only across languages but also across different grammatical constructions. Adopting a well-established experimental design, the present study examines island effects in Spanish using a speeded acceptability judgment task. To quantify variation across grammatical constructions, we tested extraction from four different types of structure (subjects, complex noun phrases, adjuncts and interrogative clauses). The results of Bayesian mixed effects modelling showed that the size of island effects varied between constructions, such that there was clear evidence of subject, adjunct and interrogative island effects, but not of complex noun phrase island effects. We also failed to find evidence that island effects were modulated by participants' working memory capacity as measured by an operation span task. To account for our results, we suggest that variability in island effects across constructions may be due to the interaction of syntactic, semantic-pragmatic and processing factors, which may affect island types differentially due to their idiosyncratic properties.

**Keywords:** extraction islands; Spanish; reading comprehension; working memory; sentence processing

## 1 Introduction

The computation of grammatical relations between non-adjacent elements is vital for successful language comprehension. For instance, in order to recover the meaning of the sentence *What did you say that Lisa baked?* comprehenders need to interpret the interrogative pronoun *what* (the *filler*) in a position different from its linear position, namely as the theme of the verb *baked* (the *gap*). Importantly, accurate comprehension involves knowing not only when to allow, but also when to disallow long-distance dependencies: some sentence regions cannot contain gaps, an observation originally made by Ross (1967), who called these regions *islands* to indicate that they disallow extraction. The attempt to establish a dependency between a filler and a gap inside an island structure generally renders a sentence unacceptable. Island effects are detectable both in participants' offline judgments and during online processing (Stowe 1986; Neville et al. 1991; Bourdages 1992; Kluender & Kutas 1993; Pickering et al. 1994; McKinnon & Osterhout 1996; Traxler & Pickering 1996; Kluender 1998; Phillips 2006; Wagers & Phillips 2009). Different structures have been found to produce island effects, including subjects (1), complex noun phrases (NPs) (2), adjuncts (3) and embedded interrogative clauses (4). In the examples below, gaps are represented by an underscore.

- (1) \***What** do you think that the speech about \_\_\_ offended Anne?
- (2) \***What** have you made the proposal that we organize \_\_\_?
- (3) \***What** did you complain when we announced \_\_\_?
- (4) \***What** did you wonder whether Lisa baked \_\_\_?

Early research attributed island effects to the violation of universal syntactic principles such as Subjacency (Chomsky 1973), the Condition on Extraction Domains (CED) (Huang 1982) or the Empty Category Principle (Chomsky 1981). However, it has long been acknowledged that the unacceptability of island sentences is subject to considerable variation, such that judgments often differ across speakers, languages and island types, among other factors. Grammatical theories have tried to account for some of this variation by postulating parametric differences between languages (Rizzi 1982), by allowing for syntactic constraints to be violable and ranked differently across languages (Legendre et al. 1995), or in terms of cumulative constraint violations resulting in different degrees of grammaticality (Haegeman et al. 2014). But it has also been suggested that syntactic constraints alone cannot capture the range of observed variation, which may be better explained from a semantic-pragmatic (e.g., Erteschik-Shir 1973; Szabolcsi & Zwarts 1993; Goldberg 2006; Truswell 2007a; Abrusán 2014) or processing perspective (Deane 1991; Kluender & Kutas 1993; Kluender 1998; 2004; Alexopoulou & Keller 2007; Hofmeister & Sag 2010). Processing-based accounts can accommodate gradience and variability by relating the severity of island effects to processing costs incurred at an island region. According to this view, the structural complexity of the sentence material intervening between filler and gap, together with semantic or pragmatic factors (such as referentiality or the presence of negation) may prevent the human parser from successfully completing the dependency, resulting in perceived ill-formedness. It has also been argued that both processing and syntactic constraints should be considered when attempting to explain this variation (e.g., Keshev & Meltzer-Asscher 2019).

To measure variation in island effects, acceptability judgments need to be collected in a systematic way and across large samples of speakers. Sprouse and colleagues created an experimental paradigm that attempts to isolate island effects from possible confounding factors such as the distance between the filler and the gap. This paradigm has been fruitfully used across languages such as English, Japanese, Brazilian Portuguese, Spanish, Italian, Norwegian, Slovenian, Hebrew and Arabic (Sprouse et al. 2011; 2012; Almeida 2014; Michel 2014; Aldosari 2015; López Sancio 2015; Sprouse et al. 2016; Kush et al. 2018; Ortega-Santos et al. 2018; Stepanov et al. 2018; Keshev & Meltzer-Asscher 2019; Kush et al. 2019; Tucker et al. 2019). The current study applies this paradigm to Spanish, a language in which there have been few empirical studies, even though previous syntactic literature has reported varying acceptability between different island types. We examine the reliability of these judgments using a large sample of Spanish native speakers. Further, we innovate on previous work by deploying a modified version of the acceptability judgment paradigm that increases processing demands by restricting the time allocated to processing sentences and providing judgments. Given previous suggestions that island effects may be modulated by individual differences in working memory (WM) capacity (see below), we also examine the influence of participants' WM capacity on their judgments. Before describing our study, we summarize previous experimental results in other languages as well as the theoretical literature on Spanish islands.

### **1.1 Cross-linguistic work on island effects**

To measure island effects, Sprouse and colleagues proposed a  $2 \times 2$  factorial design which manipulates the presence of two components of island sentences that could influence their acceptability (Sprouse et al. 2011; 2012; 2016; see (5a–d) for illustration). The

factor *distance* reflects the amount of material intervening between the filler and the gap: in the short conditions, which involve string-vacuous movement of a subject wh-pronoun, the gap immediately follows the filler in the main clause, whereas in the long conditions, the gap is more distant from the filler and located inside an embedded clause ((5a, b) vs. (5c, d)). The factor *structure* encodes the type of construction where the gap is located in the long conditions: In the non-island conditions, the gap is in a structure grammatically licensed to contain gaps, whereas in the island conditions, it is in a structure assumed to disallow them ((5a, c) vs. (5b, d)). Note that island sentences correspond to the *island/long* condition, in which the long-distance dependency is established inside a structure that disallows gaps.

- (5)
- a. *Non-island/short*: Who \_\_\_ thinks [that John bought a car]?
  - b. *Island/short*: Who \_\_\_ wonders [whether John bought a car]?
  - c. *Non-island/long*: What do you think [that John bought \_\_\_ ]?
  - d. *Island/long*: What do you wonder [whether John bought \_\_\_ ]?
- Differences-in-differences (DD) score: (5d–5b) – (5c–5a)

Crucially, because the factorial design quantifies the separate contribution of the factors *structure* and *distance*, an island effect is indicated by a *superadditive* combination, evidenced by the *island/long* condition being less acceptable than predicted by their added effects. In the view of Sprouse and colleagues (2012), this superadditivity indicates the existence of an independent grammatical constraint. Statistically, a superadditive combination should be reflected in an interaction between the factors *structure* and *distance*. Descriptively, previous studies have illustrated this interaction by calculating differences-in-differences (DD) scores on mean acceptability ratings, which are computed as shown in (5). Note, however, that superadditivity effects have also been observed in the absence of extraction from island regions and have been argued to reflect processing difficulty (Gieselmann et al. 2013; Hofmeister et al. 2014; Keshev & Meltzer-Asscher 2019). Independently of the issue of how superadditivity effects may come about, the experimental design in (5) is useful for measuring island effects whilst controlling for possible effects of distance and syntactic structure.

Sprouse and colleagues (2012) conducted one of the first experimental studies with this factorial design. In two experiments in English, they examined whether subjects, complex NPs, conditional adjuncts and polar interrogative clauses (introduced by *whether*) elicited island effects in wh-questions. They were also interested in addressing possible processing-based explanations for island effects, specifically the hypothesis that island effects might result from increased memory demands (e.g., Kluender & Kutas 1993; Kluender 1998; 2004). Under this account, the unacceptability of island sentences is not due to the violation of grammatical constraints, but rather to comprehenders' difficulty maintaining a wh-filler in memory when the local processing cost increases beyond some critical level, for example when encountering a clause boundary. To address this hypothesis, Sprouse and colleagues measured participants' WM capacity using serial recall and n-back tasks (in the latter, participants saw sequences of letters and had to press a key if one of them had been presented n-times before). The researchers hypothesized that if island effects were due to WM overload, then participants with higher WM scores should be more able to successfully parse and interpret island sentences, thus judging them as more acceptable and showing smaller DD scores (i.e., smaller island effects). However, it should be noted that predictions about how participants' WM capacity might affect the perceived severity of island violations are not so straightforward. For instance, from an incremental processing perspective, a low WM capacity might in fact make comprehenders more likely to posit a gap inside an island region because this allows for the wh-dependency to be

completed as early as possible. This might lead to an increase—rather than a decrease—in the acceptability of island violations for low WM individuals.

In Sprouse and colleagues' (2012) first experiment, participants rated the acceptability of sentences on a 7-point scale, whereas in the second experiment, ratings were collected using magnitude estimation: participants judged the experimental items by comparison to a reference sentence with a preassigned acceptability rating. Across the two experiments, interactions between *structure* and *distance* were found in all four constructions, suggesting that they all produced island effects. However, WM scores only occasionally modulated participants' DD scores (in subject islands in the first experiment and in adjunct islands in the second experiment). Furthermore, WM scores did not account for much of the variance in the data, leading Sprouse and colleagues to conclude that the interactions between *structure* and *distance* were caused by the violation of a grammatical constraint rather than processing difficulties. Subsequent studies have mostly supported the finding of a limited role of WM, as modulations have rarely been found, being restricted to either a subset of islands or dependent on a specific analysis method (Michel 2014; Aldosari 2015).

Sprouse and colleagues' paradigm was later replicated in English and extended to Japanese, Brazilian Portuguese, Spanish, Italian, Norwegian, Slovenian, Hebrew and Arabic (Sprouse et al. 2011; Almeida 2014; Michel 2014; Aldosari 2015; López Sancio 2015; Sprouse et al. 2016; Kush et al. 2018; Ortega-Santos et al. 2018; Stepanov et al. 2018; Keshev & Meltzer-Asscher 2019; Kush et al. 2019; Tucker et al. 2019).

Table 1 provides an overview of these studies, which tested either the same or a subset of the four island types examined by Sprouse and colleagues. In most cases, object extraction was tested in *wh*-questions with bare fillers (Sprouse et al. 2011; 2012; Almeida 2014; Michel 2014; Sprouse et al. 2016; Kush et al. 2018; Stepanov et al. 2018), but some studies used relative clause, left dislocation or topicalization configurations, complex fillers (e.g., *which book*), subject extraction, or resumptive pronouns (Almeida 2014; Aldosari 2015; Sprouse et al. 2016; Kush et al. 2018; Ortega-Santos et al. 2018; Keshev & Meltzer-Asscher 2019; Kush et al. 2019; Tucker et al. 2019).

Materials were similar across studies, with some exceptions: subject islands were sometimes tested with a different design than the other island types (López Sancio 2015; Sprouse et al. 2016). Further, Stepanov and colleagues (2018) and Keshev and Meltzer-Asscher (2019) tested temporal interrogatives (introduced by *when*) and Ortega-Santos and colleagues (2018) tested causal interrogatives (introduced by *why*). All studies collected acceptability judgments with untimed tasks, using either magnitude estimation (Sprouse et al. 2011; 2012; Stepanov et al. 2018) or a 7-point scale (Sprouse et al. 2012: experiment 1; Almeida 2014; Michel 2014; Aldosari 2015; López Sancio 2015; Sprouse et al. 2016; Kush et al. 2018; Ortega-Santos et al. 2018; Keshev & Meltzer-Asscher 2019; Kush et al. 2019; Tucker et al. 2019).

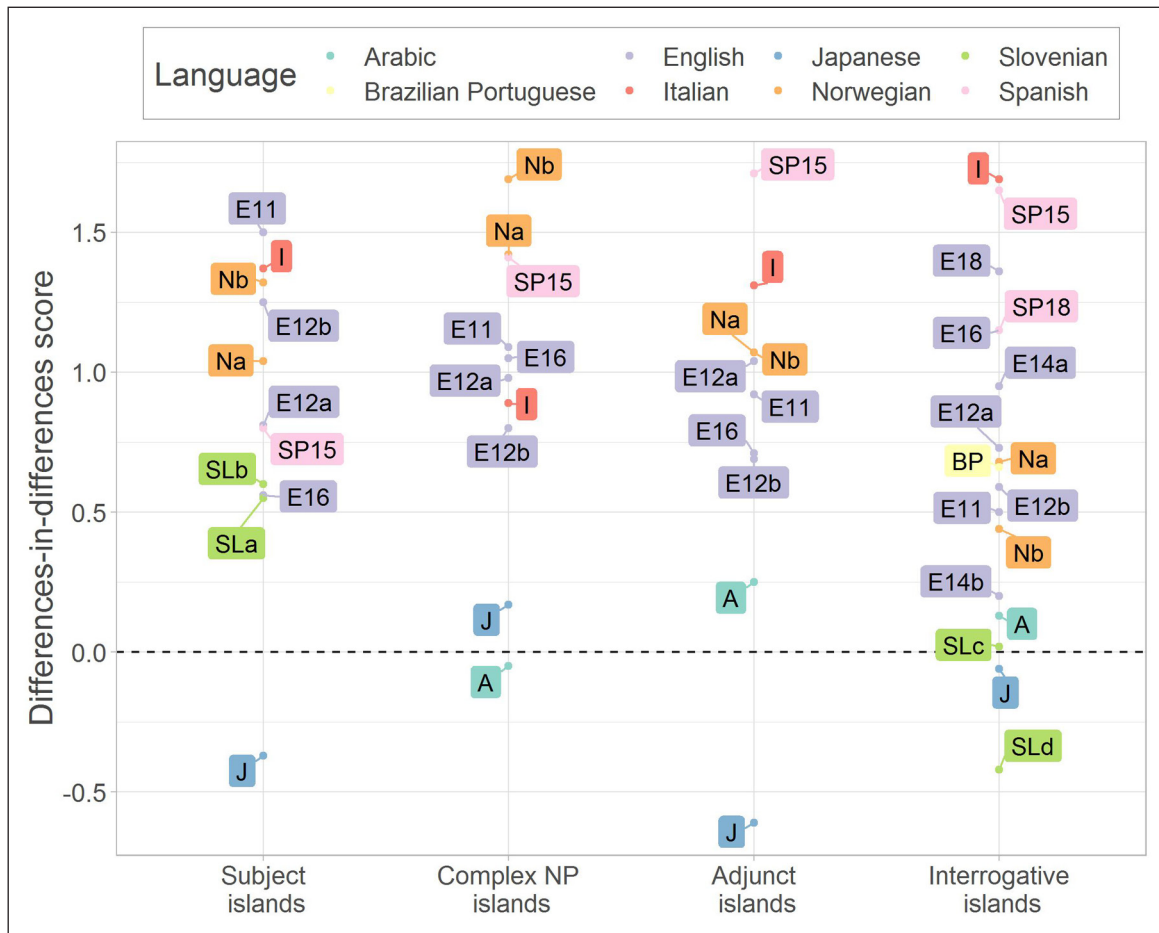
The results from previous studies are illustrated in Figure 1. Note that this figure is not a formal meta-analysis but a qualitative summary of previous findings, which we collated to identify potential generalizations about cross-linguistic variability in island effects. In line with the constructions tested in our study, Figure 1 only displays results obtained with bare fillers in *wh*-questions. We did not include studies that used complex fillers because these are influenced by additional factors that are known to affect island processing, such as discourse-linking and the lexical properties of the nouns contained in the filler (e.g., Frazier & Clifton 2002; Boxell 2014; Goodall 2015).

The distribution of DD scores in Figure 1 illustrates that there is substantial variation in the size of island effects across languages and constructions. First, although most DD scores are clearly greater than 0, consistent with superadditivity, some DD scores are

**Table 1:** Previous studies on island effects using the 2 × 2 factorial design by Sprouse and colleagues. Interrogative islands are polar unless indicated. The column *Sentence type* indicates whether island effects were tested in question, relative clause, left dislocation or topicalization configurations. *Filler* shows the type and position of the wh-filler used in question configurations: bare object (BO), complex object (CO), bare subject (BS) or complex subject (CS). *Task* shows whether judgments were collected with magnitude estimation (ME) or a 7-point scale (7-scale). Other abbreviations: Portuguese (Brazilian Portuguese), Arabic (Modern Standard Arabic), E1 (experiment 1), E2 (experiment 2).

Study	Language	Island type	Sentence type	Filler	Task
Sprouse et al. (2011)	English Japanese	Subject Complex NP Adjunct Interrogative	Question	BO	ME
Sprouse et al. (2012)	English	Subject Complex NP Adjunct Interrogative	Question	BO	7-scale (E1) ME (E2)
Almeida (2014)	Portuguese English	Interrogative	Question Left dislocation Topicalization	BO	7-scale
Michel (2014)	English	Interrogative	Question	BO	7-scale
Aldosari (2015)	English	Subject Complex NP Adjunct Interrogative	Question	CO	7-scale
López Sancio (2015)	Spanish	Subject Complex NP Adjunct Interrogative	Question Relative clause	BO	7-scale
Sprouse et al. (2016)	English Italian	Subject Complex NP Adjunct Interrogative	Question Relative clause	BO CO	7-scale
Kush et al. (2018)	Norwegian	Subject Complex NP Adjunct Interrogative Relative clause	Question	BO CO	7-scale
Ortega-Santos et al. (2018)	Spanish English	Interrogative (causal)	Question	BS	7-scale
Stepanov et al. (2018)	Slovenian	Subject Interrogative (temporal)	Question	BO BS	ME
Keshev & Meltzer-Asscher (2019)	Hebrew	Interrogative (temporal)	Relative clause	CO CS	7-scale
Kush et al. (2019)	Norwegian	Subject Complex NP Adjunct Interrogative Relative clause	Topicalization	CO	7-scale
Tucker et al. (2019)	Arabic	Complex NP Adjunct Interrogative	Question	BO CO	7-scale





**Figure 1:** Differences-in-differences (DD) scores by island type and language in wh-questions with bare fillers and without resumptive pronouns at the gap position. Different island types are shown on the x-axis and DD scores are shown on the y-axis. Correspondences are as follows: Arabic: A (Tucker et al. 2019); Brazilian Portuguese: BP (Almeida 2014); English: E11 (Sprouse et al. 2011), E12a (Sprouse et al. 2012: experiment 1), E12b (Sprouse et al. 2012: experiment 2), E14a (Almeida 2014), E14b (Michel 2014), E16 (Sprouse et al. 2016), E18 (Ortega et al. 2018); Italian: I (Sprouse et al. 2016); Japanese: J (Sprouse et al. 2011); Norwegian: Na (Kush et al. 2018: experiment 1), Nb (Kush et al. 2018: experiment 2); Slovenian: SLa (Stepanov et al. 2018: transitive verbs), SLb (Stepanov et al. 2018: unaccusative verbs), SLc (Stepanov et al. 2018: object extraction), SLd (Stepanov et al. 2018: subject extraction); Spanish: SP15 (López Sancio 2015); SP18 (Ortega et al. 2018). DD scores greater than 0 are consistent with an island effect (i.e., a super-additive effect), whereas DD scores clustered around 0 are consistent with no island effect.

below or around 0, and they were not associated with a statistically significant superadditive interaction in the original study. Notably, Arabic and Japanese do not show any reliable superadditive effects in any of the tested islands, suggesting a lack of island effects in these languages. For Japanese, the absence of island effects is perhaps attributable to the fact that wh-fillers are not displaced, in contrast with the other studies reviewed.

A second source of variability are differences between the four island types. Focusing on languages in which fillers are displaced, the distribution of DD scores seems less variable in subject, complex NP and adjunct islands than in interrogative islands, in which effects range from very large to very small or zero (e.g., compare Italian with Slovenian). Interestingly, the larger and more consistent island effects of subject, complex NP and adjunct islands is consistent with their traditional characterization as strong islands (i.e., islands that always disallow extraction), whereas the smaller and more variable interrogative island effects are compatible with their consideration as weak islands (i.e., islands that

disallow extraction selectively; Cinque 1990; Szabolcsi 2006). The mildness of interrogative island effects is also illustrated by the fact that these effects were sometimes found to be *subliminal* (Almeida 2014; Kush et al. 2018), meaning that even though these constructions yielded an interaction between *structure* and *distance*, the *island/long* sentence was still deemed acceptable or marginally acceptable.

Finally, an additional indication of variation concerns the way in which island effect sizes are ordered across languages: Figure 1 suggests that these rankings are not always consistent cross-linguistically. For instance, in Italian, interrogative constructions yielded the largest effects and complex NPs produced the smallest effects. The pattern was the opposite in Norwegian, with complex NPs yielding the largest effects and the smallest ones being caused by interrogatives. Norwegian and Italian were nevertheless similar in that adjuncts and subjects caused intermediate effects. Interestingly, the ordering of effect sizes does not only differ across languages, but also across studies in the same language. This is most clearly seen for English, the language in which island effects have most often been tested. In two of these studies, subject islands showed the largest effects, followed by complex NPs, adjuncts and interrogative clauses (Sprouse et al. 2011; 2012: experiment 2). But in Sprouse and colleagues' (2012) first experiment, DD scores were largest for adjuncts, followed by complex NPs, subjects and interrogatives, with small and graded differences between these types. Finally, in the 2016 study, interrogatives and complex NPs produced much larger DD scores than adjuncts and subjects.

To summarize, the qualitative patterns reported in previous studies show variation in island effects across languages. First, most languages do indeed exhibit island effects, but with some exceptions (Slovenian interrogative islands, Japanese and Arabic). Secondly, there are differences between island types: subjects, complex NP and adjunct constructions yield more similar island effect sizes than interrogative islands, which elicit more variability across and within languages and generally smaller effects. Finally, the ordering of island effect sizes varies across languages and sometimes even within languages. Building on these previous patterns, the current study aims to quantify differences in the acceptability of different island types within the same language, Spanish, in order to compare these with previous results. Section 1.2 describes what is known about island effects in the syntactic literature on Spanish.

### **1.2 Island effects in Spanish**

Much previous theoretical work has discussed island constraints in Spanish (Torrego 1984; Suñer 1991; Gallego & Uriagereka 2007a; b; Jiménez Fernández 2009; Gallego 2011; Haegeman et al. 2014), but this discussion has been mostly based on informal judgments, i.e., judgments based on a small number of items, typically provided by researchers themselves and/or a small number of informants, and not subject to statistical analysis. Importantly, the starting point of the theoretical work has been the assumption that island effects are universal. To our knowledge, no studies have contradicted this assumption with regard to adjuncts and complex NPs in Spanish: the literature dealing with these constructions has assumed that extraction from them is impossible, and used them as a test case to investigate other syntactic phenomena (e.g., Rivero 1978; Campos 1986; Contreras 1997; Etxepare & Uribe-Etxebarria 2005; Villa-García 2012). With regard to adjuncts, we note that there has been one previous report of apparent acceptable extractions from conditional, temporal and concessive adjuncts (Fábregas 2013). However, the author argued that these cases did not involve movement and thus, that they were no counterexamples to the generalization that extraction from adjuncts should elicit island effects in Spanish.

In contrast with adjunct and complex NPs, extraction from subjects and interrogatives appears possible in Spanish in certain contexts. With regard to subject constructions,

extraction seems to depend on whether subject phrases are located either pre- or post-verbally. But whereas some authors argue that there is a categorical ban on extraction from pre-verbal subjects (Starke 2001; Gallego & Uriagereka 2007a; b; Gallego 2011), others argue that pre-verbal subjects do not always disallow extraction (Jiménez Fernández 2009; Haegeman et al. 2014). According to these authors, the degraded acceptability of subject island sentences depends on multiple factors, which are not specific to subjects: the pre-verbality of the subject phrase, its *referentiality* (e.g., subjects with demonstrative and possessive articles are taken to be more referential than subjects with indefinite articles or quantifiers), whether the extracted element is an adjunct or an argument and whether it denotes an agent (see also Ticio 2005).

Interrogative clauses have also been thought to allow extraction in some cases. First, the question must be embedded under responsive verbs like *saber* ('to know'), which cannot take direct question quotes, rather than under rogative verbs like *preguntar* ('to ask'), which can (Torrego 1984; Suñer 1991; verb type designations are based on Lahiri 2002). Thus, (6) (taken from Torrego 1984) is claimed to be ungrammatical because the embedding verb is rogative. Object extraction (the configuration we test in our study) is subject to an additional constraint, namely, the embedded question must be of a type that does not require subject-verb inversion (Torrego 1984). Because inversion is only obligatory when the embedded question is introduced by a thematic argument of the verb, extraction should be possible out of embedded questions introduced by non-arguments such as *cómo* ('how'), *por qué* ('why') or *si* ('if'). Two examples are shown in (7) and (8), taken from Torrego (1984).

- (6) \*¿Quién pregunt-aste si — llam-ó por teléfono?  
 who ask-2SG.PST if call-3SG.PST by phone  
 'Who did you ask if — phoned?'
- (7) \*¿Qué diccionario no sab-ías a quién hab-ía devuelto —  
 what dictionary NEG know-2SG.PST to who have-3SG.PST return.PTCP  
 Celia?  
 Celia  
 'What dictionary didn't you know to whom Celia had returned?'
- (8) ¿Qué diccionario no sab-ías si Celia hab-ía devuelto —  
 what dictionary NEG know-2SG.PST if Celia have-3SG.PST return.PTCP  
 ya?  
 already  
 'What dictionary didn't you know if Celia had returned already?'

To date, we know of only two experimental studies that have used Sprouse and colleagues' design to test island effects in Spanish native speakers (López Sancio 2015; Ortega-Santos et al. 2018). Both studies elicited untimed judgments, consistently with previous studies in other languages. López Sancio (2015) tested the four types of island and found super-additive effects in all cases involving *wh*-questions (the experiment also included relative clauses constructions, which are not discussed here). These results are in line with the predictions of the theoretical literature, as interrogative islands were *if* clauses embedded under rogative verbs and subject islands were preverbal. As for Ortega-Santos and colleagues (2018), they found interrogative island effects in sentences with a responsive verb and a non-argument question word (see example (9)), against the predictions of the theoretical literature.



- (9) \*¿Quién no sab-es por qué — escrib-ió el informe?  
 who NEG know-2SG.PRS because.of what write-3SG.PST DET report  
 ‘Who don’t you know why — wrote the report?’

To summarize, most of the theoretical literature on Spanish islands has assumed that adjuncts and complex NPs disallow extraction, a prediction supported by one experimental study (López Sancio 2015). Given this previous work, we predicted these constructions to give rise to island effects in our study. By contrast, the theoretical work has reported more variable informal judgments for subject and interrogative islands, and it has also suggested that their acceptability is influenced by multiple factors, such as referentiality, syntactic position and the type of embedding verb. While López Sancio’s (2015) experiment supports these claims, Ortega-Santos and colleagues’ (2018) does not. In order to control for these variables, our materials contained those properties that were linked to unacceptability in previous work: interrogative constructions were always introduced by the verb *preguntar* (‘to ask’) and subject phrases were always definite and preverbal. Therefore, we predicted that subject and interrogative constructions should also elicit island effects, and our goal was to compare the size of these effects with the other two types of constructions, adjuncts and complex NPs.

### 1.3 The current study

Following previous work, in the current study participants were asked to judge the acceptability of experimental sentences, i.e., their naturalness as tokens of their native language. Acceptability judgments are often gathered to assess whether a sentence is grammatical or not, since grammaticality cannot be accessed directly. In addition, mentioning grammaticality in the task instructions can introduce unwanted associations with prescriptive or school-based grammar. Note that although acceptability and grammaticality are often related, they do not stand in a one-to-one relationship, as acceptability may be driven by factors other than whether a sentence obeys the rules of speakers’ mental grammar (for discussion, see Almeida 2014).

We used a speeded version of the paradigm proposed by Sprouse and colleagues (2011) to address whether Spanish speakers showed island effects during comprehension, and also whether these effects differed in strength between four different grammatical constructions: subjects, complex NPs, adjuncts and interrogative *if* clauses. A speeded acceptability task was run using a word-by-word presentation procedure. Judgments in this task have been shown to often mirror processing effects by requiring participants to rely on their WM to construct a representation of the sentence and by restricting the time available to reflect on their acceptability intuitions (Drenhaus et al. 2005; Wagers et al. 2009; Parker & Phillips 2016). We additionally examined whether participants’ WM capacity as measured by an operation span task predicted the size of island effects in participants’ acceptability judgments.

Note that the speeded task involved binary *acceptable/unacceptable* answers, in contrast with the 7-point ratings used by previous studies. The use of (simpler) binary choices was necessary because participants only had a short time (a 2 second deadline) to provide their responses. As mentioned above, short response deadlines are standardly used in speeded tasks to encourage participants to rely on their intuitions and diminish the availability of later, more strategic processes. Although previous work has suggested that binary and 7-point scale judgment tasks produce similar results, binary response tasks may be less sensitive to small contrasts (Weskott & Fanselow 2011; Schütze & Sprouse 2013). In addition, the use of binary responses required that our data were analysed with

statistical procedures in which differences between conditions are estimated in log-odds rather than percentages, in order to fulfil statistical assumptions (see below). Therefore, the use of binary responses introduced some differences between our studies and previous work, and care is necessary when comparing between them.

Our experimental predictions were as follows. Given that previous experimental and theoretical work suggests that adjunct and complex NP constructions disallow extraction in Spanish, and given that these constructions have mostly yielded strong and relatively consistent island effects in other languages, we expected them to produce clear effects in our study. As for subject and interrogative islands, different possibilities were considered. In principle, we expected them to yield clear island effects, since they contained properties that had been argued to strengthen these effects in the theoretical literature. However, given that informal judgments and previous experimental results for these island types have been less uniform, we also considered that they could produce more variation and smaller effects.

With regard to the role of WM capacity, we expected that if the unacceptability of island structures was fully or partially due to limitations in participants' WM, participants' memory scores in the operation span task should modulate island effects. Specifically, following the argumentation of Sprouse et al. (2012), we expected participants with lower memory scores to show stronger island effects compared to those with higher memory scores.

## 2 Method

### 2.1 Participants

Eighty-five native speakers of Spanish were recruited from the region of Asturias in northern Spain. Four participants were excluded because of failures in data recording, chance performance in the operation verification part of the working memory task, and/or reporting a reading impairment. Additionally, one participant was excluded due to failure to respond within the response deadline on approximately 50% of the trials of the speeded acceptability task. The remaining eighty participants had a mean age of 24 years (range: 17 – 40 years). Forty-nine participants were female and nine were left-handed. All participants provided consent, and parental consent was additionally secured for three underaged participants. To reward subjects for their participation, two 50-euro Amazon vouchers were raffled off. All procedures were in accordance with the Declaration of Helsinki.

### 2.2 Materials

Experimental items consisted of thirty-two sentence sets. There were four conditions per item and eight items per island type. Four types of construction were tested, consisting of subjects, complex NPs, adjuncts and interrogative clauses. All sentences were questions introduced by bare *wh*-fillers. An example of each island type is shown in (10) to (13).

Subject islands (10) followed the design proposed by Sprouse and colleagues (2012: Experiment 2). The extraction domain was an NP and the filler was always *who*, followed by the verb *to believe* and an embedded *that* clause. In the short conditions, the filler was associated to the subject gap position of the main verb. In both *island/short* and *non-island/short* versions, the embedded clause started with a definite subject. Island conditions differed from their non-island counterparts in that the subject contained a prepositional phrase (PP). In the *non-island/long* condition, the filler was associated to a gap that corresponded to the full embedded subject position. By contrast, the filler in the *island/long* condition was related to a gap at the PP position inside the subject phrase.

In the *island/long* condition, the preposition was fronted together with the NP due to the lack of preposition stranding in Spanish. Note that Sprouse and colleagues (2016) argued that, in languages without preposition stranding, subject islands should be tested

with an alternative design because the lack of preposition after the object may prevent participants from knowing whether the extracted element was attached to the subject or the object phrase. However, this problem did not arise in our materials because our object phrases always comprised proper names, which cannot take a prepositional phrase complement. Thus, there was no structural ambiguity in the subject constructions, such that the subject phrase was the only NP to which the prepositional phrase could be attached. In the examples below, brackets are used to mark embedded clauses.

## (10) Subject islands

a. *Non-island/short*

¿**Quién** — cre-e [que el discurso ofend-ió tanto  
 who believe-3SG.PRS COMP DET discourse offend-3SG.PST so.much  
 a Julia ayer]?  
 to Julia yesterday  
 ‘Who — believes that the discourse offended Julia so much yesterday?’

b. *Island/short*

¿**Quién** — cre-e [que el discurso de-l director  
 who believe-3SG.PRS COMP DET discourse of-DET director  
 ofend-ió a Julia]?  
 offend-3SG.PST to Julia  
 ‘Who — believes that the director’s discourse offended Julia?’

c. *Non-island/long*

¿**Quién** cre-es [que — ofend-ió tanto a Julia con  
 who believe-2SG.PRS COMP offend-3SG.PST so.much to Julia with  
 el discurso]?  
 DET discourse  
 ‘Who do you believe that — offended Julia so much with the discourse?’

d. *Island/long*

\*¿**De quién** cre-es [que el discurso — ofend-ió  
 of who believe-2SG.PRS COMP DET discourse offend-3SG.PST  
 tanto a Julia]?  
 so.much to Julia  
 ‘Of who do you believe that the discourse — offended Julia so much?’

In order to create an island structure, the modifier *of the director* was added to (10b), resulting in the ungrammatical extraction (i.e. *of who*) in the *island/long* condition (10d). Note that the four conditions—in (10) as well as the other island constructions below—sometimes differ in the presence of adverbial adjuncts: e.g., *so much* and *yesterday* in (10a). These modifiers were introduced in the non-island conditions in order to have a comparable number of words in island and non-island sentences. This was important because sentence length has been shown to modulate acceptability ratings (Konieczny 2000; Lau et al. 2017). However, whereas the experimental conditions had a comparable word number, they still differed in other dimensions (e.g., in the number of syllables or the syntactic complexity of the constructions).

The other island types conformed to the following pattern: in the long conditions, the filler was the word *qué* (‘what’), which was related to the object position of an embedded verb; in the short conditions, the filler was the word *quién* (‘who’), associated to the subject gap of the main verb. In the complex NP constructions (11), the main verb was always followed by an embedded *that* clause in the non-island conditions and by a complex NP structure in the island sentences. The complex NP structure consisted of a definite determiner followed by a noun and a complement clause, which was introduced

by a preposition (obligatory in Spanish in this context). In all cases, the preposition was *de* ('of').

(11) Complex NP islands

a. *Non-island/short*

¿**Quién** — ha ped-ido [que resolv-amos el problema  
who have.3SG.PRS ask-PTCP COMP solve-1PL.PRS DET problem  
en la reunión]?  
in DET meeting  
'Who — has asked that we solve the problem at the meeting?'

b. *Island/short*

¿**Quién** — ha hecho la petición de [que  
who have.3SG.PRS make.PTCP DET petition of COMP  
resolv-amos el problema]?  
solve-1PL.PRS DET problem  
'Who — has made the petition that we solve the problem?'

c. *Non-island/long*

¿**Qué** ha-s ped-ido [que resolv-amos — en la reunión  
what have-2SG.PRS ask-PTCP COMP solve-1PL.PRS in DET meeting  
el viernes]?  
DET Friday  
'What have you asked that we solve — at the meeting on Friday?'

d. *Island/long*

\*¿**Qué** ha-s hecho la petición de [que resolv-amos  
what have-2SG.PRS make.PTCP DET petition of COMP solve-1PL.PRS  
— el viernes]?  
DET Friday  
'What have you made the petition that we solve — on Friday?'

In the adjunct islands (12), the main verb *pensar* ('to think') was followed by a complement *that* clause in the non-island conditions. In the island conditions, the verb *protestar* ('to complain') introduced a temporal adjunct clause, headed by the word *when*.

(12) Adjunct islands

a. *Non-island/short*

¿**Quién** — pens-ó [que anunc-iamos los despido-s en  
who think-3SG.PST COMP announce-1PL.PST DET dismissal-PL in  
la reunión]?  
DET meeting  
'Who — thought that we announced the dismissals at the meeting?'

b. *Island/short*

¿**Quién** — protest-ó [cuando anunc-iamos los despido-s  
who complain-3SG.PST when announce-1PL.PST DET dismissal-PL  
en la reunión]?  
in DET meeting  
'Who — complained when we announced the dismissals at the meeting?'

c. *Non-island/long*

¿**Qué** pens-aste [que anunc-iamos — en la reunión de  
what think-2SG.PST COMP announce-1PL.PST in DET meeting of  
improviso]?  
unexpected  
'What did you think that we unexpectedly announced — at the meeting?'

d. *Island/long*

\*¿**Qué** protest-aste [cuando anunc-íamos — en la reunión  
 what complain-2SG.PST when announce-1PL.PST in DET meeting  
 de imprevisto]?  
 of unexpected  
 ‘What did you complain when we unexpectedly announced — at the meeting?’

In the interrogative constructions (13), the main verb *pensar* (‘to think’) was always used to introduce an embedded *that* clause in the non-island conditions, and the main verb *preguntar* (‘to ask’) was used to introduce an embedded *if* clause in the island conditions.

## (13) Interrogative islands

a. *Non-island/short*

¿**Quién** — piens-a [que hemos encontr-ado unas llave-s]?  
 who think-3SG.PRS COMP have.1PL.PRS find-PTCP some key-PL  
 ‘Who — thinks that we have found some keys?’

b. *Island/short*

¿**Quién** — pregunt-a [si hemos encontr-ado unas llave-s]?  
 who ask-3SG.PRS if have.1PL.PRS find-PTCP some key-PL  
 ‘Who — asks if we have found some keys?’

c. *Non-island/long*

¿**Qué** piens-as [que hemos encontr-ado — por  
 what think-2SG.PRS COMP have.1PL.PRS find-PTCP by  
 casualidad]?  
 chance  
 ‘What do you think that we have found — by chance?’

d. *Island/long*

\*¿**Qué** pregunt-as [si hemos encontr-ado — por casualidad]?  
 what ask-2SG.PRS if have.1PL.PRS find-PTCP by chance  
 ‘What do you ask if we have found — by chance?’

Materials were similar to previous studies, but some changes were made to reduce complexity and make them suitable for a speeded acceptability task. Lexical variation in the main verbs within each island type was kept to a minimum and only differences that were required by the manipulations were allowed. For instance, the words for *if* and *that* cannot be introduced by the same verb, and thus the island and non-island conditions necessarily contained different main verbs in interrogative constructions. Further, we avoided using overt pronouns or reflexives so as not to incur any additional cost of coreference processing.

### 2.3 Procedure

The experiment was conducted on two laptop PCs. Participants were tested in a quiet room at either a private residence or at the Department of Spanish Philology at the University of Oviedo in Spain. They completed a demographic questionnaire and then read the instructions of the speeded acceptability judgment task, which contained examples of acceptable and unacceptable sentences, as well as explicit instructions to not base their judgments on prescriptive/school grammar, plausibility or sentence length. After the instructions, participants were given four practice trials and the opportunity to ask questions before beginning the task.

The acceptability judgment task was run on IbeFarm (Drummond 2013), a web-based tool for collecting psycholinguistic data. Participants saw sentences word by word and indicated whether they were acceptable by pressing the keys *f* for *no* or *j* for *yes*.



Before each sentence began, a cross appeared in the center of the screen. Then, the cross was substituted by the first word of the sentence. Each word was shown for 400 ms. Once the last word disappeared, the question *Is the sentence acceptable?* appeared on screen, together with the options for *no*, presented on the left, and *yes*, presented on the right. Participants had 2000 ms to provide an answer. If no answer was given within 2000 ms, they were shown the message *Too slow* and instructed to advance to the next trial.

Experimental items were intermixed with 48 fillers and 24 items from a separate experiment (not reported here). Fillers ensured a 1:1 ratio of acceptable to unacceptable sentences and of questions to declarative sentences. The order of presentation of filler and experimental trials was pseudo-randomized on a by-participant basis, such that sentences from the same experimental condition never appeared consecutively. Experimental items were distributed across four Latin square lists, such that each participant only saw one condition of each experimental sentence. The task took about 15 minutes.

After providing acceptability judgments, participants performed an operation span task to measure their WM capacity (Turner & Engle 1989; see also Aldosari 2015). This task requires participants to recall series of items while solving mathematical operations. We adapted our version of the task from von der Malsburg (2015). The mathematical operations appeared as equations, for instance  $(2 + 7) \times 5 = 45$ . Participants had to read the equations out loud and indicate whether they were correctly solved by pressing *f* for *no* and *j* for *yes*. After each equation, a single consonant was shown on screen for 1000 ms. Each sequence of equation and consonant occurred three to five times depending on the trial. At the end of these occurrences, participants were prompted to write the letters they remembered in the order in which they had occurred. There were fifteen experimental trials, presented in a randomized order. Before the memory task, there was a pretest in which participants were asked to verify fifteen mathematical operations, without remembering any letters. The pretest was presented as a practice and participants received feedback after each operation. However, the main purpose of this part (hidden from participants), was to measure the mean time each participant spent verifying an operation in order to establish a personalized deadline (the mean response time + 2.5 standard deviations) for the following task. This was meant to ensure that fast equation solvers did not have extra time to rehearse the consonants before pressing the yes/no buttons. The operation span task took about 20 – 25 minutes. An entire experimental session lasted 35 – 45 minutes.

## 2.4 Analysis

The operation span task was scored with the partial credit unit method recommended by Conway and colleagues (2005). This means that credit was given when two or more of the consonants of a sequence were recalled in the correct order, regardless of the length of the sequence (e.g., a participant received a credit of 50% whether they recalled 1 out of 2 items or 2 out of 4 items). These by-participant WM scores were used as predictors in the analysis of the acceptability task.

In the acceptability task, the critical dependent measure was the acceptability responses given in each trial, which were coded as 0 (*unacceptable*) and 1 (*acceptable*). We did not analyse response latencies because due to the nature of our design, the target answer for three of the four conditions was *acceptable*, whereas the target answer for the *island/long* condition was *unacceptable*. This was problematic for two reasons: first, affirmative and negative responses elicit different response times (Ratcliff 1985); second, as acceptable and unacceptable responses were given with different hands, by-condition differences in response times were confounded with laterality: longer response latencies to island violations could be due either to processing disruptions or due to a right-hand advantage (most participants were right-handed).

Acceptability responses were analysed with mixed-effects logistic regression. Logistic regression is indicated for the analysis of binomial data (e.g., *acceptable* vs. *unacceptable* responses). This is because it is not appropriate to analyse binomial data by computing proportions or percentages and fitting linear models, as the distribution of proportions violates several statistical assumptions (Barr 2008; Quené & van den Bergh 2008). Specifically, proportions are inherently bounded between 0 and 1, they are not normally distributed, and their error variance is not independent from their mean. These violations can lead to biases in the statistical analyses, and consequently, to the detection of spurious effects and false null results (Jaeger 2008). By contrast, logistic regression allows probabilities of responses in each condition to be more appropriately predicted, by estimating how the different experimental manipulations change the log-odds (i.e., the log of the odds ratio) of obtaining one vs. the other type of response. Because the log-odds scale is unbounded and symmetric around zero (log-odds of zero correspond to a proportion of 50%), this approach overcomes the violations associated with analysing proportions directly.

The fixed effects in the models included *structure* (*non-island* vs. *island*) and *distance* (*short* vs. *long*), as well as their interaction. These effects were coded with treatment contrasts. Specifically, the effect of *distance* assessed the role of linear and structural distance in the absence of an island configuration by comparing the *non-island/long* and *non-island/short* conditions. The effect of *structure* assessed the cost associated with an island configuration in the absence of increased filler-gap distance by comparing the *island/short* and *non-island/short* conditions. Critically, the interaction between *structure* and *distance* addressed whether these two factors combined in an interactive way: the presence of a negative *structure*  $\times$  *distance* interaction shows that the acceptability of the *island/long* condition was lower than expected by the mere addition of the two factors. This is the statistical correlate of a superadditive effect (Sprouse et al. 2011).

The analysis addressed two additional questions about the interaction between *structure* and *distance*. First, we asked whether the size of this interaction differed for the different construction types (subject, complex NP, adjunct and interrogative). This question was addressed by adding to the model a three-way interaction between *construction type*, *structure* and *distance* (as well as all subordinate fixed effects, i.e., all two-way interactions were also computed). Second, we asked whether the size of the *structure*  $\times$  *distance* interaction was modulated by participants' WM scores, as would be expected if island effects result from WM limitations. This question was addressed by including a three-way interaction between the centered by-participant WM scores, *structure* and *distance* (as well as all subordinate fixed effects).

Mixed-effects models were fit in a Bayesian framework, which combines prior information (see below) with the evidence from the data in order to obtain a probability distribution over the plausible values of a parameter—the parameter's *posterior distribution*. Thus, an experimental effect can be quantified in terms of the likelihood of its possible magnitudes, which is more informative than a binary statement about whether the effect exists or not (Vasishth et al. 2018). For each effect of interest, we report the mean of its posterior distribution together with its 95% credible interval, which is the interval where the true mean effect lies with 95% probability. Note that Bayesian analyses do not provide *p*-values, in contrast with frequentists analyses. By way of comparison between frameworks, if an effect's 95% credible interval does not include 0, this effect would be considered significant in a frequentist framework.

The procedure for fitting Bayesian models and assessing their convergence followed recent recommendations by Vasishth and colleagues (2018). Random intercepts were used to capture variation across participants and items; random slopes were not included

because model comparisons showed that their inclusion did not lead to better models (i.e., a goodness of fit larger than 2 standard errors of the difference in LOOIC, a Bayesian measure of predictive accuracy; Bürkner 2017; Vasisht et al. 2018). To avoid making assumptions about possible effect sizes we used weakly informative priors: specifically, the prior for the fixed and random effects consisted of a normal distribution with a mean of 0 and a standard deviation of 10 log-odds. This means that 95% of the prior probability for each effect was within  $-20$  and  $20$  log-odds (practically 0%–100% in the percentage scale). Analyses were performed with the *brms* package in R (Bürkner 2017; R Core Team 2019).

Section 3 reports the estimates for the effects of theoretical interest. Readers interested in the other parameters estimated by the models, which were not of theoretical interest, can find tables with complete model outputs in the Supplementary Files. Data and materials are publicly available in the Open Science Framework repository (<https://osf.io/ckxaw/>).

### 3 Results

In the operation span task, the mean recall score was 66% (range: 27%–93%). Participants' accuracy in verifying the operations ranged from 67%–100%, with a mean of 92%. In the acceptability task, average filler accuracy was 90% (range: 73%–100%), showing that participants were able to perform adequately under speeded conditions.

The mean acceptability percentages by experimental condition are shown in Table 2. Note that differences between conditions may look different in percentages and in the log-odds scale. As noted above, binominal data should not be analysed statistically in terms of percentages because their distribution violates several statistical assumptions (Barr 2008; Quené & van den Bergh 2008). Thus, we focus our discussion on the results obtained from the statistical model that estimated differences between log-odds, but provide percentages in Table 2 for comparison.

We first assessed the effects of interest collapsing across the four constructions. The model showed clear evidence of a superadditive effect: the mean of the posterior distribution of the *structure*  $\times$  *distance* interaction was  $-3.59$  log-odds with a 95% credible interval of  $[-4.26, -2.92]$ . This negative estimate reflects the fact that the acceptability of the *island/long* condition was lower than predicted by the separate contributions of *structure* and *distance*. The effect of *distance* was  $-0.81$   $[-1.32, -0.36]$ , consistent with lower acceptability in the long than short distance conditions in non-island configurations. There was little indication of an effect of *structure* (*non-island/short* vs. *island/short* conditions), with a posterior mean of  $0.49$  and a credible interval that spanned both positive and negative values  $[-0.10, 1.07]$ . Finally, there was also little evidence that participants' WM scores modulated the *structure*  $\times$  *distance* interaction: the posterior mean of the three-way interaction was  $2.27$  and its credible interval was consistent with either a negative or positive effect  $[-1.43, 5.95]$ .

**Table 2:** Percentage of acceptances by condition and island type. Standard deviations are shown in parentheses.

	<b>Non-island short</b>	<b>Island short</b>	<b>Non-island long</b>	<b>Island long</b>
Subject	69 (46)	88 (32)	90 (30)	12 (33)
Complex NP	99 (11)	89 (32)	85 (36)	18 (39)
Adjunct	77 (42)	94 (23)	63 (48)	31 (46)
Interrogative	94 (23)	98 (14)	86 (35)	42 (50)
Overall mean	85 (36)	92 (27)	81 (39)	26 (44)

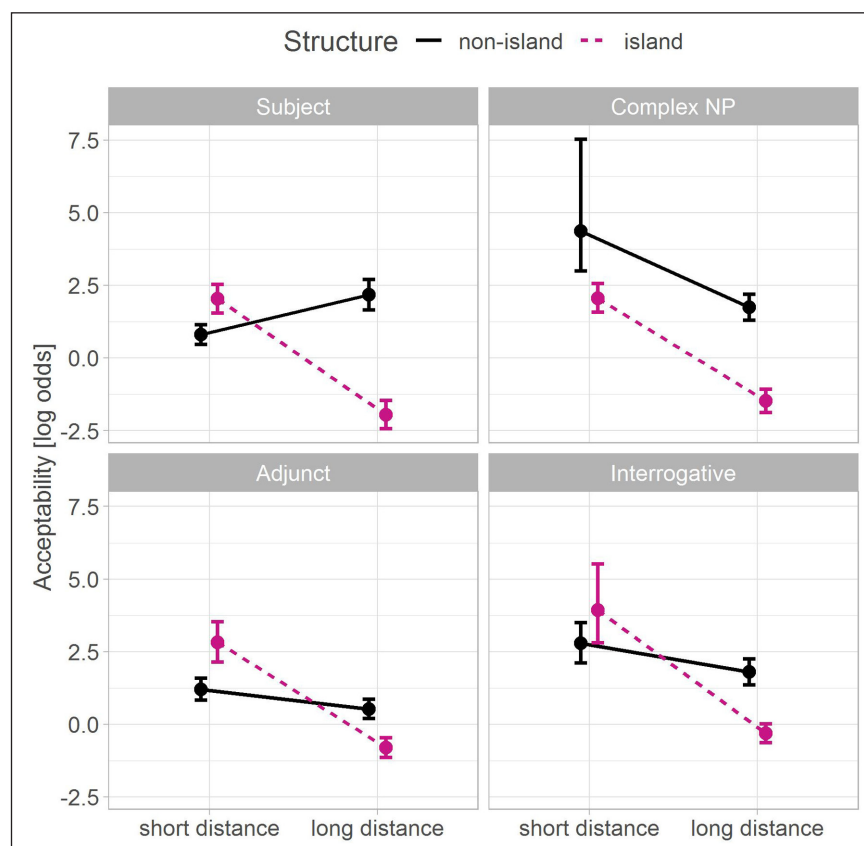
Secondly, we estimated the effects of interest for each construction separately. Figure 2 shows the acceptability patterns for each construction type (calculated in log-odds, in order to match the statistical analysis), and Figure 3 shows the posterior distributions for the *structure*  $\times$  *distance* interaction. These posterior distributions reflect the probability of the different effect sizes of the interaction given the data and the statistical model. As shown in Figure 3, the estimates of the *structure*  $\times$  *distance* interaction were largest for subject constructions, with adjunct and interrogative constructions showing intermediate effect sizes and complex NPs showing the smallest effect sizes.

### 3.1 Subject islands

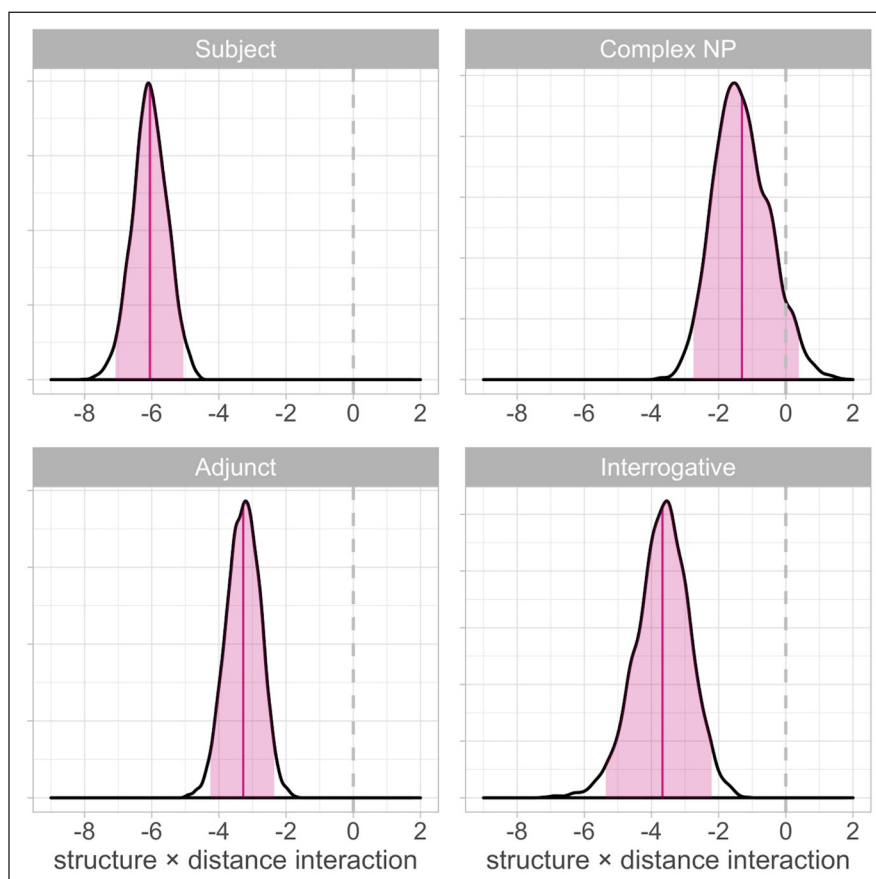
Subject constructions showed clear evidence of a *structure*  $\times$  *distance* interaction: the mean of the posterior distribution was  $-6.04$  log-odds with a credible interval of  $[-7.12, -5.08]$ . Estimates for the three-way interactions showed that the size of the *structure*  $\times$  *distance* interaction was larger for subject than for complex NP ( $4.84$  [ $3.03, 6.98$ ]), adjunct ( $2.73$  [ $1.29, 4.14$ ]) and interrogative constructions ( $2.32$  [ $0.40, 4.14$ ]). The effect of *distance* was  $1.50$  [ $0.88, 2.19$ ] and the effect of *structure* was  $1.35$  [ $0.73, 1.99$ ]. There was little indication of a *structure*  $\times$  *distance*  $\times$  *WM* interaction ( $2.23$  [ $-1.45, 6.04$ ]).

### 3.2 Complex NP islands

Surprisingly, complex NP constructions did not show clear evidence of a *structure*  $\times$  *distance* interaction: the mean of the posterior distribution was  $-1.36$  with a credible interval of  $[-2.77, 0.36]$ . Thus, although the acceptability of the *island/long* condition was lower



**Figure 2:** Acceptability by construction type. Acceptability is shown in log-odds (log-odds of zero correspond to 50% acceptability). Circles show mean acceptability by condition and error bars indicate 95% binomial confidence intervals. Non-parallel lines are indicative of an interaction between the factors *structure* and *distance*.



**Figure 3:** Posterior distributions of the *structure*  $\times$  *distance* interaction by construction type. The distributions display the posterior probability of the effect sizes of the interaction (in log-odds), with vertical bars corresponding to posterior means and shaded areas to 95% credible intervals. The dashed line corresponds to an effect size of zero.

than expected under an additive model, the size of the interaction was small and compatible with both negative and, to a smaller extent, positive values (Figure 3). Further, the interaction was smaller than for all other constructions, including subject constructions (see above), adjuncts ( $-1.95$  [ $-3.84$ ,  $-0.24$ ]) and interrogatives ( $-2.33$  [ $-4.69$ ,  $-0.26$ ]). The effect of *distance* was  $-2.77$  [ $-4.37$ ,  $-1.54$ ] and the effect of *structure* was  $-2.39$  [ $-4.00$ ,  $-1.12$ ]. Finally, there was little evidence of a *structure*  $\times$  *distance*  $\times$  *WM* interaction ( $2.23$  [ $-1.49$ ,  $6.00$ ]).

### 3.3 Adjunct islands

Adjunct constructions showed clear evidence of a *structure*  $\times$  *distance* interaction ( $-3.27$  [ $-4.28$ ,  $-2.32$ ]). This effect was smaller than in subject constructions, larger than in complex NP constructions (see above) and did not differ from interrogative constructions ( $-0.46$  [ $-2.36$ ,  $1.27$ ]). The effect of *distance* was  $-0.82$  [ $-1.36$ ,  $-0.29$ ] and the effect of *structure* was  $1.75$  [ $0.99$ ,  $2.60$ ]. As with the other constructions, there was little indication of a *structure*  $\times$  *distance*  $\times$  *WM* interaction ( $2.31$  [ $-1.36$ ,  $5.93$ ]).

### 3.4 Interrogative islands

Interrogative constructions showed clear evidence of a *structure*  $\times$  *distance* interaction ( $-3.65$  [ $-5.26$ ,  $-2.19$ ]). As shown above, this effect was smaller than in subject constructions, larger than in complex NP constructions and did not differ from adjunct constructions. The effect of *distance* was  $-1.17$  [ $-2.09$ ,  $-0.32$ ] and the effect of *structure* was  $1.21$  [ $-0.10$ ,  $2.74$ ]. As with the other constructions, there was little evidence of a *structure*  $\times$  *distance*  $\times$  *WM* interaction ( $2.31$  [ $-1.40$ ,  $6.08$ ]).



## 4 General Discussion

This study used a speeded acceptability judgment task to investigate whether Spanish native speakers showed island effects during comprehension, and whether the strength of the effects differed across subject, complex NP, adjunct and interrogative constructions. To our knowledge this is also the first attempt to use a Bayesian framework to model speeded acceptability data. Overall, the results of all construction types analysed together showed a reliable superadditive effect: island sentences were less acceptable than predicted by the combined presence of a long-distance dependency and an island structure. Our results add to the body of experimental evidence supporting the existence of island effects cross-linguistically. Furthermore, the analyses of the different constructions revealed that *structure*  $\times$  *distance* interactions were attested in three out of the four constructions, namely subject, interrogative and adjunct islands. We note that interrogative clauses and subjects were previously proposed in the Spanish theoretical literature to not show island effects under conditions different from the ones we tested, namely with responsive verbs and post-verbal, non-referential subjects (Torrego 1984; Suñer 1991; Jiménez Fernández 2009; Haegeman et al. 2014). Thus, future work should investigate whether our results extend to these contexts as well.

Regarding the role of working memory, we predicted that island effects should be stronger for participants with low WM capacity, at least under accounts that attribute island effects to a WM capacity overload (e.g., Kluender & Kutas 1993; Kluender 1998; 2004). However, our results did not support this prediction as higher operation span scores were not related to a smaller interaction between *structure* and *distance*. In fact, the interaction was larger for high-span participants, which might be expected if participants with lower WM spans are more likely than high-span participants to link the wh-filler to a gap inside an island region in order to minimise dependency length. Still, the credible intervals of these effects included both negative and positive numbers and thus there was no strong evidence in favour of this hypothesis, either. The absence of WM effects was particularly surprising because, in contrast with previous work, we tried to maximize the likelihood of WM involvement by eliciting judgments with a speeded task that is known to reflect processing effects (Drenhaus et al. 2005; Wagers et al. 2009; Parker & Phillips 2016). Therefore, although failing to find an effect is not evidence that the effect does not exist, we think that our results suggest that trying to relate WM span scores to sentence judgments is not a promising way to support memory capacity-based accounts of acceptability.

One possible reason as to why participants' operation span scores failed to modulate the effects in our data might lie in the nature of the memory mechanisms involved in processing filler-gap dependencies. Our prediction regarding a potential WM modulation of island effects (as well as our choice of WM test) was based on the assumptions of capacity models about the role of memory during language processing (e.g., Just & Carpenter 1992). However, capacity-based models have more recently been superseded by processing models that view comprehension as involving cue-based memory retrieval (e.g., McElree et al. 2003; Lewis et al. 2006), and it has been proposed that some types of island effects could be accounted for in this framework (Ortega-Santos 2011; Atkinson et al. 2016; Villata et al. 2016). From the point of view of these models, WM scores that reflect the ability to recall serial order information might not necessarily correlate with comprehension difficulty (Gieselmann et al. 2013; see also Van Dyke et al. 2014). Alternatively, the lack of WM effects could be related to the use of binary judgments, which may vary less between participants than judgments on a scale, thus reducing the likelihood of finding a WM modulation (Schütze & Sprouse 2013).

With regard to the comparison between different grammatical constructions, we found important differences in the size of the interactions between *structure* and *distance*, which

was the largest for subject islands, the smallest for complex NP islands and had intermediate size for interrogative and adjunct islands. This may be unexpected from the perspective of grammatical theories that reduce island effects to the violation of a single underlying principle (Chomsky 1973; Sabel 2002; Müller 2010; 2011): if this were the case, then more uniform island effects would have been expected across constructions. Rather, our results indicate that not all island effects have the same cause, consistent with accounts that do not treat islands as a natural class, either by differentiating between weak and strong islands (Cinque 1990; Szabolcsi 2006) or by making distinctions within those classes (Stepanov 2007). However, our findings do not fully match any of the previously proposed class distinctions.

Our finding of differences between island types is consistent with previous work in other languages, but the ordering of island effect sizes across constructions was unexpected. Recall that we had made two alternative predictions. One possibility was for all island types to yield strong effects, since they all contained properties that have been associated with islandhood. Another possibility was that subject and interrogative islands might yield weaker effects than adjunct and complex NP islands, because more variability and cases of acceptable extraction had been reported for the latter in previous work on Spanish. Our results did not meet any of these predictions, since subject island violations actually yielded the largest superadditive effect, complex NPs produced the smallest and interrogative and adjunct clauses showed intermediate effect sizes. Interestingly, López Sancio (2015) also found that interrogative and adjunct clauses yielded greater effect sizes than complex NPs, although his results differ from ours in that subject islands yielded the weakest effect. In what follows, we will examine the factors that may be responsible for the differential size of island effects in Spanish, focusing on by-construction idiosyncratic properties and on the differences between effects in percentages and in the log-odds scale used to analyse the acceptability data.

#### 4.1 Subject islands

Subject islands yielded the strongest superadditivity effects. Although extraction from embedded subjects violates both the CED constraint (which prohibits extraction from non-complements) and Subjacency (on the assumption that both DP and CP are bounding nodes in Spanish; Torrego 1984), previous work had suggested that subject islands might only yield weak effects. Previous research also indicated that the effects might increase if subjects are pre-verbal or referential, and if the extracted element is an adjunct or denotes an agent (Starke 2001; Ticio 2005; Gallego & Uriagereka 2007a; b; Jiménez Fernández 2009; Gallego 2011; Haegeman et al. 2014). The *island/long* condition in the subject island configuration, which obtained the lowest rating of all island types, did indeed have these properties: subjects were pre-verbal and referential, since they were introduced by a definite article, and the extracted element could be considered an adjunct or an argument of the NP denoting an agent. Definiteness is well known to be an islandhood-inducing cue (e.g., Fiengo & Higginbotham 1981). Thus, our results support the prediction that subject island effects are strong when the island sentences bear these characteristics.

We also note that, while this strong superadditive effect is compatible with previous research, it might appear bigger in our results than subject island effects really are in Spanish for two reasons. The first is that the *non-island/long* condition was accepted more often than its short counterpart. This pattern seems related to the rating of the *non-island/short* condition, which was low relative to the other grammatical conditions. The cause of these differences is unclear. One possible explanation is that the NPs that we used (e.g., *the discourse*) subcategorize a PP agent (Escandell Vidal 1995; Lorenzo González 1995) and that the NP sounded less natural without them in the

*non-island/short* condition. Note that the same NPs were used without a PP in the *non-island/long* condition, but they were less discourse-prominent and the filler of the sentence denoted the agent, so presumably the lack of the PP did not affect them in the same way. Future research could avoid NPs that might subcategorize a PP to bypass potential confounds.

The second possible reason for the large superadditivity effect is that the independent effect of extracting a PP from an NP was not controlled for (i.e., it was only tested in the *island/long* condition and thus it was not factored out). As a consequence, any potential decrease in acceptability associated with it is included in the superadditivity measure (Kush et al. 2018). Note that this problem does not arise in the other construction types because they do not involve sub-extraction from an NP. This is a caveat for the comparison of island effect sizes across constructions that future research should address.

Our results also differ to those of López Sancio (2015), who found that subject islands yielded the weakest effect of all four island types. This might be explained by the fact that López Sancio used indefinite subjects and an alternative design for subject islands that may underestimate the size of the effect (see Kush et al. 2018 for discussion). The contrast between our results and López Sancio's underlines the importance of using the same design across studies to facilitate cross-linguistic comparison.

#### 4.2 Complex NP islands

Complex NPs yielded very weak island effects, despite our prediction that they would show a strong superadditive pattern, and despite wh-extraction also crossing two bounding nodes here. Note that the *island/long* condition was still in the unacceptable range (below 0 in the log-odd scale), and so, the weak effect should not be taken as an indication that complex NP island sentences are acceptable in Spanish. However, our results suggest that their unacceptability is not strongly determined by the fact that a wh-dependency is being established inside the complex NP, contrary to syntactic accounts of island effects. Instead, they indicate that the low acceptability is caused by the mere presence of the complex NP, independently of the fact that a filler-gap dependency is resolved inside it. This is because the *island/short* condition elicited lower acceptability than the *non-island/short* condition, a pattern unique to the complex NP construction. The NP structure in the island conditions may have reduced acceptability relative to its non-island counterpart due to its greater structural complexity: island conditions consisted of an NP and a prepositional complement, which in turn contained an embedded clause (e.g., lit. *the petition of that we solve the problem*). By contrast, non-island conditions only had an embedded clause in the corresponding structural position, which depended directly on the verb (e.g., *that we solve the problem*). Therefore, the greater structural complexity of the island conditions, together with the presence of an additional discourse referent (the NP head) may have increased processing demands.

Furthermore, we note that the size of the superadditivity effect is much greater in percentages than in log-odds: the mean acceptability of the *island/long* condition on the percentage scale was below 25%, suggesting that this condition was strongly unacceptable for most of speakers. However, this pattern changes when judgments are expressed in the log-odds scale. This is because in logistic regression, the estimated effects relate to the predicted probabilities via a non-linear sigmoid (i.e., S-shaped) function, meaning that small differences in percentages may be much larger in log-odds when they take place at the extremes of the percentage scale (closer to 0% and 100%) than when they are closer to the middle point of 50%. This impacts the interaction sizes of the complex NP materials due to their extremely high mean acceptability percentage in the *non-island/short* condition (99%). It is beyond the scope of this paper to discuss which scale most appropriately

reflects acceptability, a question that exceeds the domain of island effects. Here we used the log-odds scale because it would have been statistically inappropriate to analyse percentages with a linear model. However, we acknowledge that the log-odds scale may not reflect the intuitive acceptability measures provided in the syntactic literature, and we think that this is an issue that deserves further research and more discussion between syntacticians and psycholinguists who work with acceptability judgments.

### 4.3 Adjunct islands

Adjunct constructions only yielded moderate island effects, and the *island/long* condition had relatively high acceptability. Given that extraction from adjuncts violates the CED constraint and that is generally thought to yield strong and mostly consistent island effects across languages (Stepanov 2007; see also Figure 1), and since they are also assumed to disallow extraction in the literature on Spanish, we might have expected them to yield a larger superadditive pattern.

Our results suggest that extraction from adjuncts in Spanish is less unacceptable than predicted by traditional syntactic accounts, because syntactic constraints are generally taken to impose a categorical ban on extraction. The increased acceptability and reduced island effect could be related to the possibility of obtaining a single event reading of the whole sentence, which has been previously argued to ameliorate extraction from adjuncts (Truswell 2007a; b; 2011; Müller 2017). Specifically, the single event reading could be obtained by interpreting the event expressed by the temporal *when* adjunct clause as the cause of the event denoted by the main verb. We think that this interpretation was favoured in our items for two reasons. First, all temporal clauses were embedded under the verb *to complain*, which denotes an event that is generally a reaction to something (i.e., it has a cause). Second, the *when* clauses contained expressions such as *too soon* or *without permission*, which may have facilitated the interpretation that the denoted event was inappropriate and thus a plausible reason for complaining.

From an incremental processing perspective, analysing the fronted wh-pronoun *qué* ('what') as the direct object of the embedded verb requires overcoming the potential intervention effect caused by having crossed another operator (*cuando* 'when'). Participants who were able to parse our adjunct island sentences successfully might then have considered them acceptable. Note also that in our adjunct island items only three words intervened between the filler and the gap (as compared to four in subject and interrogative islands and seven in complex NPs). This might have reduced processing difficulty, maximizing the likelihood that the filler remains activated until the gap is encountered or can be successfully retrieved from memory at this point. In addition, the filler is unlikely to be initially mistaken for the object of the main verb *protestar* ('to complain'), because this is almost always intransitive.

### 4.4 Interrogative islands

Interrogative islands also yielded intermediate superadditivity effects. Previous research indicated that, when there were differences between island types, interrogatives typically produced smaller interactions, but it also suggested that a rogative embedding verb (i.e., a verb that can take a direct question quote) could strengthen the effects (Suñer 1991). All of our items contained one such verb, *preguntar* ('to ask'), and thus the strong interaction might be related to its presence. Future research should test directly whether rogative verbs are less permeable to extraction than responsive verbs (which do not take direct question quotes), and whether the effect is restricted to Spanish or affects other languages as well. Note that, if rogative and responsive verbs impose different constraints on extrac-



tion, some of the variation in the size of interrogative island effects across previous studies (Figure 1) could stem from the choice of embedding verb.

While this strong effect might be consistent with the claim that rogative verbs such as *to ask* increase interrogative island effects, it must be noted that, just like in our complex NP conditions, the interaction size in percentages differed from the interaction size in log-odds: in percentages, interrogative islands yielded the weakest effect of all island types (42%; see Table 2), suggesting that extraction from interrogative islands was not considered strongly unacceptable. Further research could address this possibility by gathering scalar rather than binary judgments of Spanish interrogative constructions.

Note that from a processing perspective we would not have expected any dramatic differences in acceptability between adjunct and interrogative island violations. Our interrogative island sentences were structurally similar to our adjunct island sentences, with movement of an interrogative object pronoun crossing another operator located at the embedded clause boundary but with relatively few words and no referential noun phrases intervening between filler and gap.

#### **4.5 Concluding remarks**

Our study found experimental evidence for the existence of adjunct, subject and interrogative island effects in Spanish, thus supporting the cross-linguistic generality of such effects. However, our study also found differences in the size of superadditive effects across constructions, indicating that not all islands are equally strong and that these acceptability differences are not always predictable by the principles proposed in the theoretical work. For example, although both subject islands and adjunct islands violate the CED constraint, the two types of violation differed considerably in their acceptability and the size of the superadditivity effect. Extraction from complex NPs yielded only weak superadditivity effects despite violating Subjacency. This challenges accounts that attempt to reduce all island effects to a single grammatical principle (Chomsky 1973; Sabel 2002; Müller 2010; 2011) or to a set of such principles (Cinque 1990; Szabolcsi 2006). Instead, it suggests that the perceived severity of island effects is influenced by construction-specific properties such as structural complexity or the presence of intervening operators, as well as by factors such as referentiality and the choice of matrix verb or subordinating conjunction.

The observed superadditivity effects are also compatible with processing-based accounts of island effects (Gieselmann et al. 2013). However, the debates as to whether island effects reflect properties of the grammar or the workings of the parser seems to us to be rather futile. This is because some of the traditionally invoked syntactic principles (such as Subjacency) could also be seen as formalisations of processing constraints (Kluender & Kutas 1993). A more promising way forward might be to try to capture variation in island effects within constraint-based frameworks that allow for individual constraints—regardless of their nature or origin—to be violable and to differ in their relative weightings (e.g., the Gradient Symbolic Computation framework; Smolensky et al. 2014). In this way, we might be able to account for island effects in terms of interacting constraints which may possibly be universal, but which might be differently weighted across languages (similar in spirit to Legendre et al.'s 1995 Optimality Theoretic approach).

On the methodological side, we hope to have illustrated the usefulness of the Bayesian framework for analysing acceptability judgment data. But our results also highlight another potentially important methodological issue: we observed some discrepancies between participants' mean acceptability ratings and the transformed rating scores used for the statistical analysis. We hope that these findings trigger further discussion as to how best to deal with binary acceptability data.



## Additional File

Additional files for this article, including materials, data and analysis code, can be found at the Open Science Framework website:

- **Supplementary files.** <https://osf.io/ckxaw/>. DOI: <https://doi.org/10.5334/gjgl.1058.s1>

## Abbreviations

1 = first person, 2 = second person, 3 = third person, CED = Condition on Extraction Domains, COMP = complementizer, CrI = credible interval, DD = differences-in-differences, DET = determiner, ms = milliseconds, NEG = negation, NP = noun phrase, PL = plural, PP = prepositional phrase, PRS = present, PST = past, PTCP = participle, SG = singular, WM = working memory

## Acknowledgements

CP's work in this article was partially supported by a Severo Ochoa predoctoral grant of the Government of Asturias, Spain (PA-17-PF-BP16105), a grant from Fundació Banco Sabadell and a grant of the Spanish Government (FEDER/Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación) to project DaLiV (FFI2017-87699-P). We thank Guillermo Lorenzo, Dave Kush and Bruno Nicenboim for useful comments and guidance, and Jon Sprouse, Diogo Almeida and Sergio López Sancio for kindly sharing their data and work.

## Competing Interests

The authors have no competing interests to declare.

## Author Contributions

CP, SL and CF designed the experiment and CP and EV ran it. CP, SL and JV analysed the data and all authors contributed to the writing of the paper.

## References

- Abrusán, Marta. 2014. *Weak island semantics*. New York, NY: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199639380.001.0001>
- Aldosari, Saad Mohammed. 2015. *The Role of individual differences in the acceptability of island violations in native and non-native speakers*. Lawrence, KS: University of Kansas dissertation.
- Alexopoulou, Theodora & Frank Keller. 2007. Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language* 83(1). 110–160. DOI: <https://doi.org/10.1353/lan.2007.0001>
- Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Abrallin* 13(2). 55–93. DOI: <https://doi.org/10.5380/rabl.v13i2.39611>
- Atkinson, Emily, Aaron Apple, Kyle Rawlins & Akira Omaki. 2016. Similarity of wh-phrases and acceptability variation in wh-islands. *Frontiers in Psychology* 6. 2048.1–16. DOI: <https://doi.org/10.3389/fpsyg.2015.02048>
- Barr, Dale J. 2008. Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language* 59(4). 457–474. DOI: <https://doi.org/10.1016/j.jml.2007.09.002>
- Bourdages, Johanne S. 1992. Parsing Complex NPs in French. In Helen Goodluck & Michael Rochemont (eds.), *Island constraints: Theory, acquisition, and processing*, 61–87. Dordrecht: Springer. DOI: [https://doi.org/10.1007/978-94-017-1980-3\\_3](https://doi.org/10.1007/978-94-017-1980-3_3)

- Boxell, Oliver. 2014. Lexical fillers permit real-time gap-search inside island domains. *Journal of Cognitive Science* 15(1). 97–136. DOI: <https://doi.org/10.17791/jcs.2014.15.1.97>
- Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. DOI: <https://doi.org/10.18637/jss.v080.i01>
- Campos, Héctor. 1986. Indefinite object drop. *Linguistic Inquiry* 17(2). 354–359.
- Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232–286. Holt: Rinehart and Winston.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110884166>
- Cinque, Guglielmo. 1990. *Types of A'-dependencies*. Cambridge, MA: MIT Press.
- Contreras, Heles. 1997. Algunas observaciones sobre la subyacencia [Some observation on subadjacency]. In Marianna Pool Westgaard & Sergio Bogard (eds.), *Estudios de lingüística formal*, 199–209. Mexico: Colegio de México. DOI: <https://doi.org/10.2307/j.ctv47w5ck.13>
- Conway, Andrew R.A., Michael J. Kane, Michael F. Bunting, D. Zach Hambrick, Oliver Wilhelm & Randall W. Engle. 2005. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12(5). 769–786. DOI: <https://doi.org/10.3758/BF03196772>
- Deane, Paul. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2(1). 1–64. DOI: <https://doi.org/10.1515/cogl.1991.2.1.1>
- Drenhaus, Heiner, Stefan Frisch & Douglas Saddy. 2005. Processing negative polarity items: When negation comes through the backdoor. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 145–164. Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110197549.145>
- Drummond, Alex. 2013. Ibex Farm. <http://spellout.net/ibexfarm/>.
- Erteschik-Shir, Nomi. 1973. *On the nature of island constraints*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Escandell Vidal, María Victoria. 1995. *Los complementos del nombre* [Noun complements]. Madrid: Arco Libros.
- Etxepare, Ricardo & Myriam Uribe-Etxebarria. 2005. In-situ wh-phrases in Spanish: Locality and quantification. *Recherches Linguistiques de Vincennes* 33. 9–34. DOI: <https://doi.org/10.4000/rlv.1238>
- Fábregas, Antonio. 2013. Nota sobre unas construcciones que temblaría la gramática si fueran extracciones de isla [A note on some constructions that grammar would tremble if they were island extractions]. *Signo y Señal* 24. 175–188.
- Fiengo, Robert & James Higginbotham. 1981. Opacity in NP. *Linguistic Analysis* 7. 395–421.
- Frazier, Lyn & Charles Jr. Clifton. 2002. Processing “d-linked” phrases. *Journal of Psycholinguistic Research* 31(6). 633–659. DOI: <https://doi.org/10.1023/A:1021269122049>
- Gallego, Ángel J. 2011. Successive cyclicity, phases, and CED effects. *Studia Linguistica* 65(1). 32–69. DOI: <https://doi.org/10.1111/j.1467-9582.2010.01175.x>
- Gallego, Ángel J. & Juan Uriagereka. 2007a. Conditions on sub-extraction. In Luis Eguren & Olga Fernández Soriano (eds.), *Coreference, modality, and focus: Studies on the syntax-semantics interface*, 45–70. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.111.04gal>
- Gallego, Ángel J. & Juan Uriagereka. 2007b. Sub-extraction from subjects: A phase theory account. In José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Deprez & María José Cabrera (eds.), *Romance linguistics 2006: Selected papers from the 36th*

- linguistic symposium on Romance languages (LSRL)*, 155–168. Philadelphia, PA: John Benjamins. DOI: <https://doi.org/10.1075/cilt.287.12gal>
- Gieselmann, Simone, Robert Kluender & Ivano Caponigro. 2013. Isolating processing factors in negative island contexts. In Yelena Fainleib, Nicholas LaCara & Yangsook Park (eds.), *Proceedings of NELS 41*. 233–246. Amherst, MA: Graduate Linguistic Student Association.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- Goodall, Grant. 2015. The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology* 5. 1493.1–11. DOI: <https://doi.org/10.3389/fpsyg.2014.01493>
- Haegeman, Liliane, Ángel L. Jiménez-Fernández & Andrew Radford. 2014. Deconstructing the Subject Condition in terms of cumulative constraint violation. *The Linguistic Review* 31(1). 73–150. DOI: <https://doi.org/10.1515/tlr-2013-0022>
- Hofmeister, Philip & Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86(2). 366–415. DOI: <https://doi.org/10.1353/lan.0.0223>
- Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2014. Processing effects in linguistic judgment data: (Super-) additivity and reading span scores. *Language and Cognition* 6. 111–145. DOI: <https://doi.org/10.1017/langcog.2013.7>
- Huang, Cheng-Teh James. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. DOI: <https://doi.org/10.1016/j.jml.2007.11.007>
- Jiménez Fernández, Ángel Luis. 2009. On the composite nature of subject islands: A phase-based approach. *SKY Journal of Linguistics* 22. 91–138.
- Just, Marcel Adam & Patricia A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99(1). 122–149. DOI: <https://doi.org/10.1037/0033-295X.99.1.122>
- Keshev, Maayan & Aya Meltzer-Asscher. 2019. A processing-based account of subliminal wh-island effects. *Natural Language and Linguistic Theory* 37(2). 621–657. DOI: <https://doi.org/10.1007/s11049-018-9416-1>
- Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. *Syntax and Semantics* 29. 241–280. DOI: [https://doi.org/10.1163/9789004373167\\_010](https://doi.org/10.1163/9789004373167_010)
- Kluender, Robert. 2004. Are subject islands subject to a processing account? In Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez & Benjamin Schmeiser (eds.), *Proceedings of WCCFL 23*. 101–125. Somerville, MA: Cascadilla Press.
- Kluender, Robert & Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8(4). 573–633. DOI: <https://doi.org/10.1080/01690969308407588>
- Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29(6). 627–645. DOI: <https://doi.org/10.1023/A:1026528912821>
- Kush, Dave, Terje Lohndal & Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language and Linguistic Theory* 36(3). 743–779. DOI: <https://doi.org/10.1007/s11049-017-9390-z>
- Kush, Dave, Terje Lohndal & Jon Sprouse. 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>

- Lahiri, Utpal. 2002. *Questions and answers in embedded contexts*. Oxford: Oxford University Press.
- Lau, Jey Han, Alexander Clark & Shalom Lappin. 2017. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science* 41(5). 1202–1241. DOI: <https://doi.org/10.1111/cogs.12414>
- Legendre, Geraldine, Colin Wilson, Paul Smolensky, Kristin Homer & William Raymond. 1995. Optimality and wh-extraction. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers 18), 607–636. Amherst, MA: GLSA. DOI: <https://doi.org/10.7282/T31Z463D>
- Lewis, Richard L., Shrvan Vasishth & Julie A. Van Dyke. 2006. Computation principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10. 447–454. DOI: <https://doi.org/10.1016/j.tics.2006.08.007>
- López Sancio, Sergio. 2015. *Testing syntactic islands in Spanish*. Vitoria-Gasteiz: Universidad del País Vasco/Euskal Herriko Unibertsitatea master thesis.
- Lorenzo González, Guillermo. 1995. *Geometría de las estructuras nominales: Sintaxis y semántica del SDet* [Geometry of nominal structures: Syntax and semantics of DP]. Oviedo: Universidad de Oviedo.
- McElree, Brian, Stephani Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48. 67–91. DOI: [https://doi.org/10.1016/S0749-596X\(02\)00515-6](https://doi.org/10.1016/S0749-596X(02)00515-6)
- McKinnon, Richard & Lee Osterhout. 1996. Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials. *Language and Cognitive Processes* 11(5). 495–524. DOI: <https://doi.org/10.1080/016909696387132>
- Michel, Dan. 2014. *Individual cognitive measures and working memory accounts of syntactic island phenomena*. San Diego, CA: University of California dissertation.
- Müller, Christiane. 2017. Extraction from adjunct islands in Swedish. *Norsk Lingvistisk Tidsskrift* 35(1). 67–85.
- 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>
- Müller, Gereon. 2011. *Constraints on displacement: A phase-based approach*. Amsterdam: John Benjamins.
- Neville, Helen, Janet L. Nicol, Andrew Barss, Kenneth I. Forster & Merrill F. Garrett. 1991. Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3(2). 151–165. DOI: <https://doi.org/10.1162/jocn.1991.3.2.151>
- Ortega-Santos, Iván. 2011. On Relativized Minimality, memory and cue-based parsing. *Iberia: International Journal of Theoretical Linguistics* 3(1). 35–64.
- Ortega-Santos, Iván, Lara Reglero & Jon Franco. 2018. Wh-Islands in L2 Spanish and L2 English: Between poverty of the stimulus and data assessment. *Fontes Linguae Vasconum* 126. 435–471.
- Parker, Dan & Colin Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition* 157. 321–339. DOI: <https://doi.org/10.1016/j.cognition.2016.08.016>
- Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82(4). 795–823. DOI: <https://doi.org/10.1353/lan.2006.0217>
- Pickering, Martin, Stephen Barton & Richard Shillcock. 1994. Unbounded dependencies, island constraints and processing complexity. In Charles Jr. Clifton, Lyn Frazier & Keith Rayner (eds.), *Perspectives on sentence processing*, 199–224. Hillsdale, NJ: Lawrence Erlbaum.



- Quené, Hugo & Huub van den Bergh. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59(4). 413–425. DOI: <https://doi.org/10.1016/j.jml.2008.02.002>
- Ratcliff, Roger. 1985. Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review* 92(2). 212–225. DOI: <https://doi.org/10.1037/0033-295X.92.2.212>
- R Core Team. 2019. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rivero, María-Luisa. 1978. Topicalization and wh movement in Spanish. *Linguistic Inquiry* 9(3). 513–517.
- Rizzi, Luigi. 1982. Violations of the wh-island constraint and the Subjacency condition. In *Issues in Italian syntax*, 49–76. Dordrecht: Foris.
- Ross, John Robert. 1967. *Constraints on variables in syntax*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- 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>
- Schütze, Carson T. & Jon Sprouse. 2013. Judgment data. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139013734>
- Smolensky, Paul, Matthew Goldrick & Donald Mathis. 2014. Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science* 38(6). 1102–1138. DOI: <https://doi.org/10.1111/cogs.12047>
- Sprouse, Jon, Ivano Caponigro, Ciro Greco & Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34(1). 307–344. DOI: <https://doi.org/10.1007/s11049-015-9286-8>
- Sprouse, Jon, Matt Wagers & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. DOI: <https://doi.org/10.1353/lan.2012.0004>
- Sprouse, Jon, Shin Fukuda, Hajime Ono & Robert Kluender. 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>
- Starke, Michal. 2001. *Move dissolves into merge: A theory of locality*. Geneva: University of Geneva dissertation.
- Stepanov, Arthur. 2007. The End of CED? Minimalism and extraction domains. *Syntax* 10(1). 80–126. DOI: <https://doi.org/10.1111/j.1467-9612.2007.00094.x>
- Stepanov, Arthur, Manca Mušič & Penka Stateva. 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>
- Stowe, Laurie A. 1986. Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1(3). 227–245. DOI: <https://doi.org/10.1080/01690968608407062>
- Suñer, Margarita. 1991. Indirect questions and the structure of CP: Some consequences. In Héctor Campos & Fernando Martínez-Gil (eds.), *Current studies in Spanish linguistics*, 283–312. Washington, DC: Georgetown University Press.
- Szabolcsi, Anna. 2006. Strong vs. weak islands. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, 479–531. Malden, MA: Blackwell. DOI: <https://doi.org/10.1002/9780470996591.ch64>



- Szabolcsi, Anna & Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1(3). 235–284. DOI: <https://doi.org/10.1007/BF00263545>
- Ticio, Emma M. 2005. Locality and anti-locality in Spanish DPs. *Syntax* 8(3). 229–286. DOI: <https://doi.org/10.1111/j.1467-9612.2005.00080.x>
- Torrego, Esther. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15(1). 103–129.
- Traxler, Matthew & Martin Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35(3). 454–475. DOI: <https://doi.org/10.1006/jmla.1996.0025>
- Truswell, Robert. 2007a. Extraction from adjuncts and the structure of events. *Lingua* 117(8). 1355–1377. DOI: <https://doi.org/10.1016/j.lingua.2006.06.003>
- Truswell, Robert. 2007b. Tense, events, and extraction from adjuncts. In Malcolm Elliott, James Kirby, Osamu Sawada, Eleni Staraki & Suwon Yoon (eds.), *Proceedings of CLS* 43. 233–247. Chicago, IL: Chicago Linguistic Society.
- Truswell, Robert. 2011. *Events, phrases, and questions*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199577774.001.0001>
- Tucker, Matthew A., Ali Idrissi, Jon Sprouse & Diogo Almeida. 2019. Resumption ameliorates different islands differentially: Acceptability data from Modern Standard Arabic. In Amel Khalfaoui & Matthew Tucker (eds.), *Perspectives on Arabic linguistics* 30. 1–52. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/sal.7.09tuc>
- Turner, Marilyn L. & Randall W. Engle. 1989. Is working memory capacity task dependent? *Journal of Memory and Language* 28(2). 127–154. DOI: [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- Van Dyke, Julie A., Clinton L. Johns & Anuenue Kukona. 2014. Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition* 131. 373–403. DOI: <https://doi.org/10.1016/j.cognition.2014.01.007>
- Vasishth, Shravan, Bruno Nicenboim, Mary E. Beckman, Fangfang Li & Eun Jong Kong. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71. 147–161. DOI: <https://doi.org/10.1016/j.wocn.2018.07.008>
- Villa-García, Julio. 2012. *The Spanish complementizer system: Consequences for the syntax of dislocations and subjects, locality of movement, and clausal structure*. Storrs, CT: University of Connecticut dissertation.
- Villata, Sandra, Luigi Rizzi & Julie Franck. 2016. Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua* 179. 76–96. DOI: <https://doi.org/10.1016/j.lingua.2016.03.004>
- von der Malsburg, Titus. 2015. Py-Span-Task – A software for testing working memory span. <https://github.com/tmalsburg/py-span-task>.
- Wagers, Matthew W. & Colin Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45. 395–433. DOI: <https://doi.org/10.1017/S0022226709005726>
- Wagers, Matthew W., Ellen F. Lau & Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61(2). 206–237. DOI: <https://doi.org/10.1016/j.jml.2009.04.002>
- Weskott, Thomas & Gisbert Fanselow. 2011. On the informativity of different measures of linguistic acceptability. *Language* 87(2). 249–273. DOI: <https://doi.org/10.1353/lan.2011.0041>

**How to cite this article:** Pañeda, Claudia, Sol Lago, Elena Vares, João Veríssimo and Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa: a journal of general linguistics* 5(1): 21.1–30. DOI: <https://doi.org/10.5334/gjgl.1058>

**Submitted:** 02 August 2019

**Accepted:** 21 November 2019

**Published:** 02 March 2020

**Copyright:** © 2020 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/>.



*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS The Open Access icon, which is a stylized padlock with a circular arrow around it.