A translation-matched, experimental comparison of three types of wh-island effects in Spanish and English

Claudia Pañeda[a], Sandra Villata[b], Dave Kush[c], and Jon Sprouse[d]

[a]Universidad de Oviedo, [b]Università degli Studi di Enna "Kore", [c]University of Toronto, [d]New York University Abu Dhabi

Abstract. According to received wisdom, *wh*-extraction from embedded *wh*-questions gives rise to island effects in English, but not in Spanish. This was an important observation for the development of a parameters-based theory of cross-linguistic variation in islands. However, recent experimental work has found some *wh*-island effects in Spanish, raising questions about whether the cross-linguistic contrast holds for all speakers. We address this in 12 acceptability judgment experiments with about 100 participants per experiment and translation-equivalent materials. In each language, we examine *wh*-extraction from 3 *wh*-clause types (introduced by *whether*, *why* and *when*) under 2 matrix verb types (*know* and *ask*), amounting to 6 *wh*-islands that are relevant to assess the reported contrasts. We test (i) for the presence or absence of *wh*-island effects in the two languages, (ii) for a gradient contrast in effect size, and (iii) for evidence of increased individual variation in Spanish as compared to English. We find (i) that *wh*-island effects are present in both English and Spanish, (ii) that they are rather large in both languages and larger in Spanish for most *wh*-island types, and (iii) that Spanish does not show more individual variation than English. Our results speak against the cross-linguistic contrast as originally proposed, suggesting that its use as evidence for theories that encode cross-linguistic variation in *wh*-island effects might need to be reconsidered.

Keywords: wh-islands; cross-linguistic variation; Spanish; English; acceptability judgments

1 Introduction

In her seminal paper on subject-verb inversion in Spanish, Torrego (1984) observed that Spanish and English differ with respect to *wh*-islands: extraction from embedded *wh*-questions in English gave rise to island effects, whereas extraction from embedded *wh*-questions in Spanish appeared to be possible, at least for certain types of *wh*-questions and matrix verbs (a complexity we review in section 2). This observation was offered as evidence toward a theory of subject-verb inversion and successive cyclic movement, but it also became a critical example of cross-linguistic variation in island effects, and contributed to the development of the influential parameterized Subjacency theory, in which languages differ with respect to which phrasal categories act as bounding nodes (see also Rizzi 1982). However, recent experimental work has raised questions about whether this observation holds for all speakers of European Spanish (see section 2). Given the discrepancy between older reports and new findings, our goal is to solidly establish the empirical facts of cross-linguistic variation in *wh*-islands between Spanish and English. We report 12 acceptability judgment experiments (6 in each language) that used translation-equivalent materials to examine extraction from 3 *wh*-clause types under two matrix verb types, amounting to 6 *wh*-islands that are relevant to the extraction patterns reported in Torrego (1984). These experiments were designed to (i) test for the presence or absence of *wh*-island effects in the two languages, (ii) test for a gradient contrast in effect size, in case the difference is one of size rather than presence/absence, and (iii) test for evidence of increased individual variation in Spanish as compared to English (using both between-participant and within-participant measures), which might explain some of the variability in the observations reported in the literature. We address these questions with data from approximately 100 participants per experiment, to ensure enough statistical power to detect small differences in effect sizes, and four tokens per condition per participant, to be able to examine within-participant variation.

Anticipating slightly, our results suggest (i) that *wh*-island effects are present for *wh*-extraction in both English and Spanish, (ii) that the effects are fairly large in both languages, and perhaps unexpectedly, larger in Spanish for most *wh*-island types; and (iii) that Spanish does not show increased individual variation based on either between-participant or within-participant measures. The rest of this article is organized as follows. Section 2 discusses Torrego's (1984) observation in more detail, including the structural (and potentially lexical) conditions that are reported to be necessary to extract a *wh*-word from a *wh*-island in (European) Spanish. Section 2 also reviews recent experimental work in both Spanish and English that has raised the possibility that extraction form *wh*-islands is not available for all speakers of Spanish. Section 3 describes the logic and design of our 12 experiments in detail. Section 4 presents the results with analyses for each of our three driving questions. Section 5 discusses the consequences of our results for theories of cross-linguistic variation in island effects. Section 6 presents a brief conclusion.

2 Previous work motivating the current study

Wh-island effects are the unacceptability that arises when a phrase is A′ moved from a *wh*-clause. According to the received wisdom, there are *wh*-island effects for *wh*-extraction in English (1), but not in Spanish (2).

(1) *What did he wonder where John put?

(Chomsky 1964, p. 47)

(2) ¿Qué dices que no te explicas por qué
what say.2SG.PRS that NEG 2SG.DAT explain.2SG.PRS why

Juan se habrá comprado?
Juan 3SG.DAT have.3SG.FUT buy.PTCP

'What do you say that you can't figure out why Juan may have bought for himself?'

(Torrego 1984, p. 115)

The contrast above is based on Torrego's (1984) seminal article on subject-verb inversion, but, in fact, her observations were more nuanced: she considered extraction from *wh*-islands to be readily available from embedded subject positions, but more constrained from embedded object positions, which was only deemed possible when the *wh*-island was introduced by a non-argument, like *si* ('whether/if'), *por qué* ('why'), or *cuándo* ('when').[1] Given that there is an independent comp-trace effect for extraction from embedded subject positions in English (Perlmutter 1968; Bresnan 1977), a well-controlled cross-linguistic contrast is only expected in sentences with extraction from *wh*-islands introduced by non-arguments. Interestingly, though, recent acceptability judgment studies have found island effects in Spanish in such cases (López-Sancio 2015; Ortega-Santos et al. 2018; Pañeda et al. 2020; Rodríguez & Goodall 2020; Stigliano & Xiang 2021; Pañeda & Kush 2022). The majority of these results have been obtained with the verb *preguntar*(*se*) ('to ask/wonder') (López-Sancio 2015; Pañeda et al. 2020; Stigliano & Xiang 2021), which independently prevents extraction according to Torrego (see also Suñer 1991). But *wh*-island effects have also been observed with *saber* ('to know'), which is not claimed to pose any such constraint (Ortega-Santos et al. 2018; Pañeda & Kush 2022). Thus, previous experimental results suggest that currently spoken Spanish manifests *wh*-island effects even in the cases in which it was predicted not to do so. These data, together with repeated findings of *wh*-island effects in English (e.g., Sprouse et al. 2012; Sprouse et al. 2016), cast doubt on the cross-linguistic contrast often inferred from Torrego's (1984) observations.

Nonetheless, the conclusions that can be reached about the cross-linguistic contrast based on the experimental studies above are limited, for several reasons. First,

[1] In Torrego's analysis, extraction from these *wh*-islands is ultimately possible to the extent that they do not require subject-verb inversion. While she makes the generalization that inversion is not required in *wh*-islands introduced by non-arguments, she also points out that there could be variation depending on the non-argument. However, we set aside questions of the theory of subject-verb inversion (the focus of Torrego's paper) to focus solely only on the empirical question of variation in *wh*-island effects.

those studies have often tested extraction out of a single type of *wh*-island (usually *whether* islands under *ask*/*wonder*), and it is unclear whether their results generalize to other cases. Second, previous studies have mostly tested the two languages separately, with different materials and, sometimes, under different conditions that are not comparable. For example, Pañeda & Kush's (2022) Spanish results cannot be compared to Sprouse et al.'s (2012) English results because the former were obtained with complex (or "d(iscourse)-linked") extractees, which have been claimed to independently reduce or eliminate *wh*-island effects (Pesetsky 1987; see Sprouse & Villata 2021 for a review), while the latter were obtained with bare extractees. We are aware of one study that tested *wh*-island effects in both languages (Ortega-Santos et al. 2018), but this study only assessed *know why* islands and extraction was from the embedded subject position, which, as we indicated above, yields a comp-trace effect in English, making cross-linguistic comparisons difficult.

To better assess the cross-linguistic contrast, the current study tests *wh*-island effects in both languages in a wider range of syntactic configurations, using the same translation-equivalent lexicalizations to rule out items as a source of variation. Because Spanish and English are predicted to differ with regard to extraction from *wh*-clauses introduced by non-arguments, we test three such clauses—*whether*, *why* and *when* clauses—, and because the *know*/*ask* contrast is predicted to affect *wh*-island effects, we test them under both embedding verbs. We not only address the binary question of whether island effects are present or absent in each language, but also the gradient question of whether they differ in size across languages: for instance, *wh*-island effects could be smaller in Spanish, supporting a weaker version of the cross-linguistic claim. In addition, given the relative uncertainty in the literature regarding *wh*-islands in Spanish (with judgments perhaps varying across speakers), we investigate whether there is increased individual variation in Spanish as compared to English (using both between-participant and within-participant measures), in case that is a potential source of the discrepancy between informal observations and experimental findings.

3 The design of our study

We ran twelve acceptability judgment experiments, six in English and six in Spanish, each examining one of three *wh*-island types (*whether*, *why* or *when*) under one of two embedding verbs (*know*, *ask*). Island effects were examined in sentences with *wh*-extraction (e.g. *What did the politician ask when they would reject?*), in line with most previous studies on both languages (e.g., Sprouse et al. 2012; Ortega-Santos et al. 2018; Pañeda et al. 2020; Pham et al. 2020). Extractees were bare (e.g. *What* rather than *which cake*), given that complex or "d-linked" fillers may independently reduce *wh*-island effects (Pesetsky 1987). The same item set was used in all experiments, varying only the words of interest. Across languages, translation-equivalent items were used. Each experiment was 55 items long, consisting of 16 target items (4 tokens of each of the 4 conditions in the 2×2 factorial design for island effects, described below), 32 filler items spanning the range of acceptability, and 7 (unmarked) practice items spanning the range of acceptability. The task used a 7-point rating scale. For each experiment, we recruited 112 participants using Prolific (www.prolific.co). In the following sections we discuss the design and methods in more detail.

3.1 The factorial design for island effects

To maximize compatibility with both the theoretical and experimental literature, we adopted the 2×2 factorial design that has become standard in the island effects literature. Under this design, island effects are quantified as an interaction between two factors: STRUCTURE, which manipulates the structure of the embedded clause between a declarative (non-island) and a question (island), and POSITION, which manipulates whether the gap is in the matrix or embedded clause (see Sprouse, 2007 and subsequent work). This design controls for the independent effects on acceptability of these two factors, and isolates the island effect in the interaction term, driven by an unexpected low rating in the island/embedded condition (unexpected because it is lower than predicted by the combination of the independent effects of the sentence containing an island and having a gap in an embedded clause). Example (3) illustrates this design for Spanish, with the alternative verbs and wh-phrases separated by slashes. The English design is illustrated in the translations:

(3) a. non-island/matrix

¿Quién __ pensaba que rechazarían la propuesta?
who __ think.3SG.PST that reject.3PL.COND the.F proposal
Who __ thought that they would reject the proposal?

b. non-island/embedded

¿Qué pensaba el político que rechazarían __?
what think.3SG.PST the.M politician that reject.3PL.COND __
What did the politician think that they would reject __?

c. island/matrix

¿Quién __ preguntó / quería saber
who __ ask.3SG.PST / want.3SG.PST know
si / por qué / cuándo rechazarían la propuesta?
whether / why / when reject.3PL.COND the.F proposal
Who __ asked / wanted to know whether / why / when they would reject the proposal?

d. island/embedded

¿Qué preguntó / quería saber el político
what ask.3SG.PST / want.3SG.PST know the.M politician
si / por qué / cuándo rechazarían __?
whether / why / when reject.3PL.COND __
What did the politician ask / want to know whether / why / when they would reject __?

Island effects are identified statistically as a 'superadditive' Position × Structure interaction (we describe our statistical models in section 4). Island effects can be identified visually in a plot of the means of the four conditions as non-parallel lines arranged such that the island/embedded condition is lower than the other three conditions, as in the left and center panel of Figure 1. If the island/embedded condition has much lower acceptability than the other conditions, as in the left panel, this is an indication that there is a large island effect; if it only has slightly lower acceptability, as

in the center panel, there is a small island effect. The right panel shows that the absence of island effects can be identified visually as parallel lines.

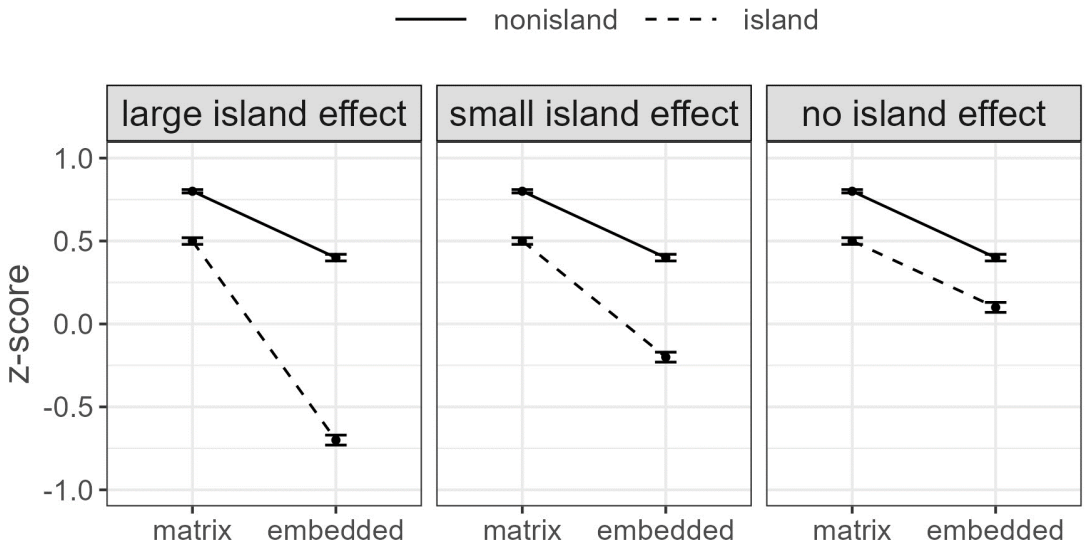

Figure 1: Example of how large, small and no island effects appear in plots of the means of the four conditions in the 2×2 design.

The size of the island effect can be quantified with a differences-in-differences (DD) score, as in (4).

(4) DD score: (non-island/embedded – island/embedded) – (non-island/matrix – island/matrix)

3.2 Participants

We recruited 1344 participants in total through Prolific: 112 for each of the 12 experiments. They were paid $2.75 USD for their participation. To identify native speakers of US English or European Spanish, we used three tools. First, we used Prolific's prescreening tools to identify individuals who considered the target language their first language and had mostly lived in either the US or Spain before turning 18. Second, we asked all participants questions about where they lived from birth until age 13, and about the languages that were spoken in their home as children. Finally, we included two trials within the experiment where we asked participants to read a short description of an ethically challenging situation, and write at least one complete sentence in the target language about how they would respond. We excluded participants from analysis if they reported not living in the target country, not speaking the target language (or speaking it as a non-dominant language) in their home, or if their responses to the morality trials appeared non-native to us. We also removed participants from analysis if they responded to 4 or more fillers with a rating that was more than 2 standard deviations away from the mean rating for that filler. The final sample sizes of each experiment are shown in Table 1.

Structure	English	Spanish
know whether	95	105
ask whether	96	104
know why	97	98
ask why	101	104
know when	99	97
ask when	101	101

Table 1: Number of participants that met the inclusion criteria and were included in the analysis for each experiment.

3.3 Materials

For each language, we created sixteen item sets based on the four conditions in the factorial design (quadruplets). These sixteen sets were used in all experiments (with the island type and verb modifications). The English and Spanish sentences were translation-equivalents and as lexically-matched as possible. We did have to make several choices to ensure a clear test of island effects. First, the embedded subject in English was always an overt pronoun (either *they* or *you*), but a null pronoun in Spanish. We chose null pronouns in Spanish because we perceived them as more natural than overt pronouns in our sentences, and because their position relative to the verb is not observable, meaning that participants can posit it as preverbal or postverbal, as they prefer. We could have alternatively used an overt determiner phrase subject in a fixed position, but we decided not to do so because Torrego (1984) observed that there may be independent factors that influence the acceptability of the subject-verb vs verb-subject orderings, which in turn may be confounded with, or even interact with, island effects. While this is an interesting question in its own right, we abstract away from it here, allowing participants to posit the most acceptable subject position in their grammar, giving us a test of island effects alone. Second, in the *know* experiments, we used *want to know* (rather than simply *know*) because *want to know* seems to highlight the interrogative nature of the complement *wh*-clause, similar to *ask*. Third, in the embedded clause, we used verbs that we perceived to be transitively-biased to facilitate the interpretation of the *wh*-word as an embedded object. Finally, we presented embedded verbs in the conditional tense (e.g., *would reject*) to make a modifier interpretation of the *wh*-questions (particularly *cuándo*/*when*-islands) less likely (such interpretation would make them an adjunct island rather than a *wh*-island). The full set of materials is available as Supplementary files (S1).

In addition to the 16 experimental item sets, we selected 32 pre-tested fillers and 7 pre-tested practice items per language that evenly span the full range of acceptability ratings. The English items were taken from Sprouse et al. (2013) and the Spanish items from Ortega-Santos (2020).

3.4 Presentation

Items were distributed across four lists using a Latin Square procedure, such that participants rated each of the 4 conditions 4 times, and each time the item was from a unique item set (no repetitions of lexical items). Participants first saw instructions with 3 example items with ratings of 1, 4, and 7 to demonstrate the task. They then rated items themselves. The first 7 items were the practice items in the same (pseudorandom) order for each participant (but not marked as practice, so they appeared as simply trials in the experiment). The next two items were two of the filler items (a very low and very

high rating). The rest of the experiment contained the 16 experimental items for that particular list and 30 remaining fillers in a pseudorandom order such that there was at least one filler between two experimental items. The experiments were run using Qualtrics (Provo, Utah). It took participants about 10 minutes to complete an experiment.

4 Results

We z-score transformed the results by participant to eliminate common forms of scale biases prior to analysis. We constructed linear mixed effects models for each of the questions of the study (described below). We then calculated two inferential statistics for the critical interaction terms: null hypothesis *p*-values using the lmerTest package (Kuznetsova et al. 2017), and Bayes factors (BF) using the BayesFactor package (Morey et al. 2022) in R (R Core Team 2023). We included BFs because they provide distinct information to *p*-values – they represent the ratio of the probability of the data under the experimental hypothesis to the probability of the data under a null hypothesis. We interpreted a *p*-value less than .05 as statistically significant, a BF greater than 3 as meaningful evidence in favor of the presence of an interaction (i.e., the data is 3x more likely under the experimental hypothesis than under the null hypothesis), and a BF less than .33 as evidence against the presence of an interaction (i.e., the data is 3x more likely under the null hypothesis than under the experimental hypothesis). To check whether our BFs were robust to the choice of priors, we calculated them with the three different priors built-in to the BayesFactor package. In the text, we only report the BFs obtained with a medium width prior, but all three widths yield equivalent results unless otherwise indicated.

4.1 The presence of island effects

Our first question is whether each of the 6 island effects is present in the two languages. Figure 2 shows the interaction plots for each *wh*-island under each verb in the two languages, along with the differences-in-differences or DD scores, an estimate of the interaction term that we calculated as in (4). To assess the presence of island effects statistically, we constructed linear mixed effects models crossing Structure × Position for each of the 12 island effects, with the maximal random effect structure that did not result in convergence failure. The presence of an island effect would show up as a significant Structure × Position interaction. Figure 2 also shows the *p*-values and Bayes factors for the interaction term in these models. The full results are shown in Table 2 for English and Table 3 for Spanish.

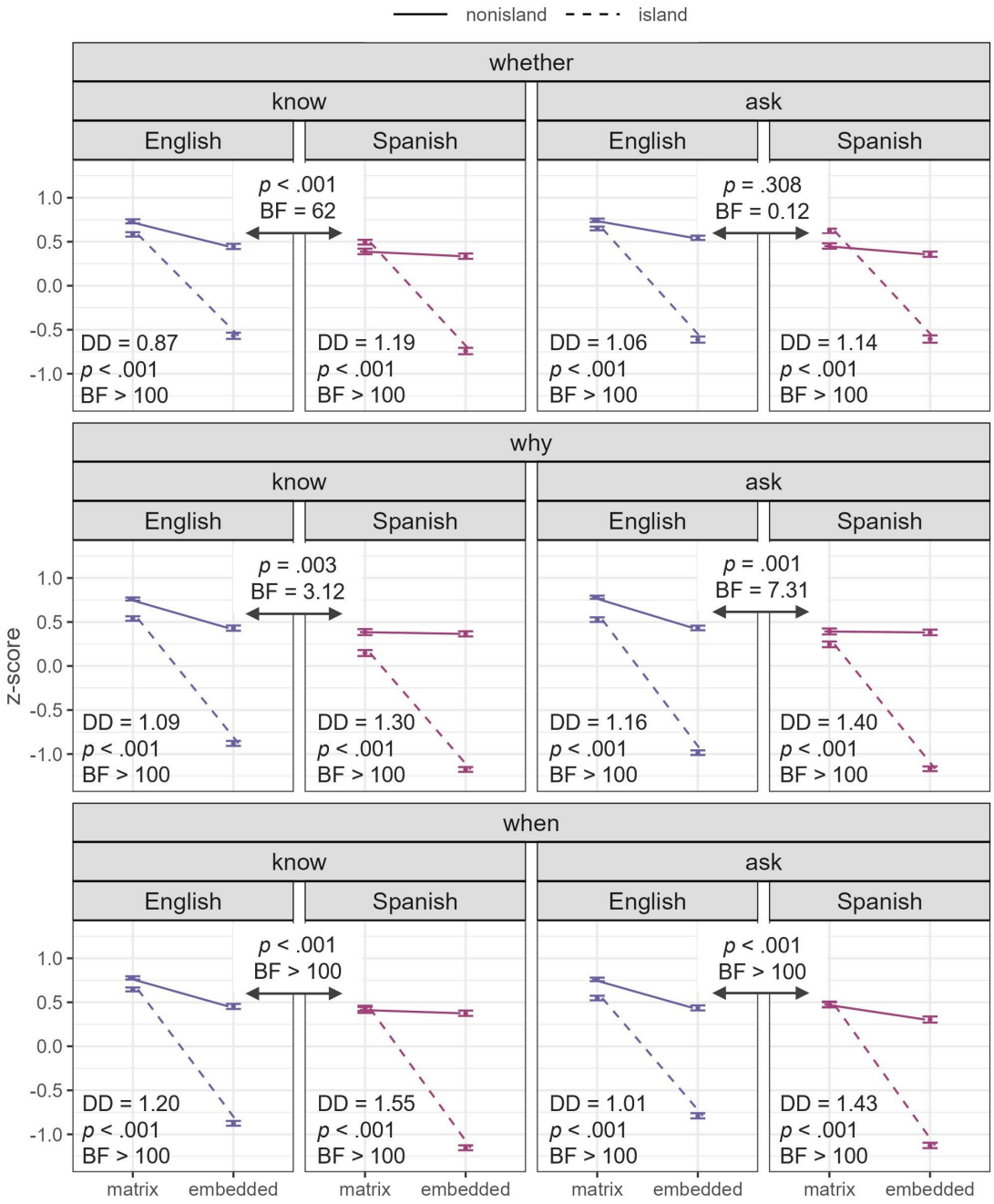

Figure 2: Interaction plots for *whether*, *why* and *when* islands under both *know* and *ask* in English and Spanish. For each island in each language, we show the differences-in-differences (DD) score, the *p*-value of the superadditive Structure × Position interaction indicating an island effect and the Bayes Factor (BF). Above the arrows, we also show the *p*-value and the BF for the three-way Structure × Position × Language interactions, which assess cross-linguistic differences in size (see section 4.2).

	Estimate	SE	t	*p*	BF
Know whether					
Intercept	0.299	0.028	10.710	< .001	
Structure	−0.291	0.019	−15.480	< .001	
Position	−0.359	0.020	−17.830	< .001	
Structure × Position	−0.217	0.018	−12.380	< .001	> 100
Ask whether					
Intercept	0.332	0.026	12.910	< .001	
Structure	−0.312	0.017	−17.980	< .001	
Position	−0.365	0.021	−17.410	< .001	
Structure × Position	−0.266	0.016	−16.880	< .001	> 100
Know why					
Intercept	0.213	0.024	8.800	< .001	
Structure	−0.383	0.020	−18.780	< .001	
Position	−0.438	0.020	−22.400	< .001	
Structure × Position	−0.272	0.021	−13.190	< .001	> 100
Ask why					
Intercept	0.189	0.019	10.100	< .001	
Structure	−0.418	0.018	−22.760	< .001	
Position	−0.465	0.021	−22.560	< .001	
Structure × Position	−0.291	0.018	−15.870	< .001	> 100
Know when					
Intercept	0.249	0.023	10.970	< .001	
Structure	−0.364	0.020	−18.220	< .001	
Position	−0.462	0.018	−25.620	< .001	
Structure × Position	−0.299	0.021	−14.440	< .001	> 100
Ask when					
Intercept	0.240	0.026	9.161	< .001	
Structure	−0.359	0.025	−14.555	< .001	
Position	−0.414	0.023	−18.393	< .001	
Structure × Position	−0.251	0.021	−11.892	< .001	> 100

Table 2: Results of the Structure × Position linear mixed models run on each of the English experiments and Bayes Factors (BF) for the Structure × Position interactions. The factors followed an effects coding scheme: Structure (non-island: −1, island: 1), Position (matrix: −1, embedded: 1). All models included random intercepts and Structure and Position slopes for participant and item. The Structure × Position interaction was included in the slopes whenever this converged.

	Estimate	SE	t	*p*	BF
Know whether					
Intercept	0.120	0.038	3.137	.004	
Structure	−0.242	0.021	−11.563	< .001	
Position	−0.324	0.019	−16.637	< .001	> 100
Structure × Position	−0.296	0.014	−21.548	< .001	
Ask whether					
Intercept	0.207	0.043	4.819	< .001	
Structure	−0.200	0.025	−7.866	< .001	
Position	−0.331	0.022	−14.750	< .001	> 100
Structure × Position	−0.285	0.019	−15.344	< .001	
Know why					
Intercept	−0.069	0.031	−2.209	.038	
Structure	−0.445	0.024	−18.302	< .001	
Position	−0.336	0.030	−11.339	< .001	> 100
Structure × Position	−0.326	0.016	−20.245	< .001	
Ask why					
Intercept	−0.036	0.034	−1.060	.301	
Structure	−0.425	0.027	−16.010	< .001	
Position	−0.357	0.025	−14.050	< .001	> 100
Structure × Position	−0.350	0.020	−17.410	< .001	
Know when					
Intercept	0.017	0.036	0.466	.645	
Structure	−0.376	0.024	−15.577	< .001	
Position	−0.406	0.021	−19.027	< .001	> 100
Structure × Position	−0.388	0.023	−16.677	< .001	
Ask when					
Intercept	0.032	0.031	1.021	.318	
Structure	−0.359	0.027	−13.381	< .001	
Position	−0.444	0.024	−18.353	< .001	
Structure × Position	−0.355	0.014	−25.330	< .001	> 100

Table 3: Results of the Structure × Position linear mixed models run on each of the Spanish experiments and Bayes Factors (BF) for the Structure × Position interactions. The factors followed an effects coding scheme: Structure (non-island: −1, island: 1), Position (matrix: −1, embedded: 1). All models included random intercepts and Structure and Position slopes for participant and item (except for *know whether*, where only a Structure slope was included for item to ensure convergence). The Structure × Position interaction was included in the slopes whenever this converged.

In all cases, the island/embedded condition was much less acceptable than the other three conditions, resulting in the visual patterns indicative of superadditive Structure × Position interactions, which were confirmed statistically (all $p < .001$ and all BFs > 100). This suggests that all 6 island effects are present in both languages.

4.2 The size of island effects

Though there does not appear to be any cross-linguistic variation in the presence of island effects (a binary question), there could still be variation in the size of the island effects (a gradient question). For example, *wh*-island effects could be present but smaller in Spanish. To test this possibility, we constructed linear mixed effects models for each of the 6 island types, but crucially combined the results of the two languages. These models crossed Structure × Position × Language and included the maximal random effect structure that converged. A difference in island effect size between the two languages would show up as a significant three-way interaction. A summary of the three-way interaction effects is shown in Table 4. The full results of the models are available as Supplementary files (S2). The *p*-values and Bayes factors of the critical three-way interaction term are also shown in Figure 2 above the arrows.

	Estimate	SE	t	*p*	BF
Know whether	−0.040	0.009	−4.259	< .001	62
Ask whether	−0.009	0.009	−1.020	.308	0.12
Know why	−0.027	0.009	−3.018	.003	3.12
Ask why	−0.030	0.009	−3.325	.001	7.31
Know when	−0.044	0.009	−5.174	< .001	> 100
Ask when	−0.052	0.009	−5.693	< .001	> 100

Table 4: Summary of the critical Structure × Position × Language interaction effects obtained in the six Structure × Position × Language linear mixed models run on each island and verb combination. Bayes Factors (BF) for the interaction are also provided.

We find a statistically significant size difference for 5 out of the 6 island types by both null hypothesis testing ($p < .05$) and Bayes factor (BF > 3): *know whether*, *know when*, *know why*, *ask why*, and *ask when*. From these, all BFs are robust to prior widths except for *know why*, where the BF is inconclusive with a wide (BF = 1.65) and an ultrawide prior (BF = 1.56). As for *ask whether*, it is not significant by *p*-value ($p = .308$) and shows evidence against a size difference by BF (BF = .12). However, it is important to note that the direction of the size differences we observed is opposite to the one that might be expected under a weaker version of the original cross-linguistic claim – Spanish island effects are larger than English island effects by about .3 z-units on average. This suggests that while there is evidence of a gradient form of cross-linguistic variation for 4 or 5 out of 6 island types, it is not in line with the original observation.

4.3 Individual variation

Our third question is whether there is more individual variation in Spanish than English. We ask this because there can be individual variation, representing different idiolects, that is obscured by focusing on sample means. Idiolectal variation could explain the apparent discrepancy between our results and Torrego's (1984) observations. In other words, even if, overall, we find island effects, some speakers of Spanish may manifest no or smaller island effects while others show large island effects (i.e., between-participant variation). It is also possible that some speakers of Spanish show more variability within their own judgments than English speakers (i.e., within-participant variation), suggesting that they may have two grammars at their disposal. To be clear, we expect some amount of variation both between- and within-participants. That is the nature of behavioral studies (they are inherently noisy). The critical question is whether Spanish shows more variation than English in one or both of these ways.

To look for between-participant variation, we plot two sets of distributions: the island effect sizes calculated as by-participant DD scores (Figure 3) and the by-participant z-score means of the island/embedded conditions (Figure 4; the non-island/embedded conditions are also shown as a control; see Kush et al. 2018; 2019; Bondevik et al. 2021 for similar analyses). We can look for a visual pattern suggesting two or more populations of speakers, which would appear as a bimodal (or multimodal) distribution. We see no obvious sign of bimodality in the Spanish experiments, either in the DD scores or in the island/embedded conditions. Some signs of bimodality are instead observable in the English experiments (e.g. the *ask whether* DD scores and the *ask when* DD scores and island/embedded conditions).

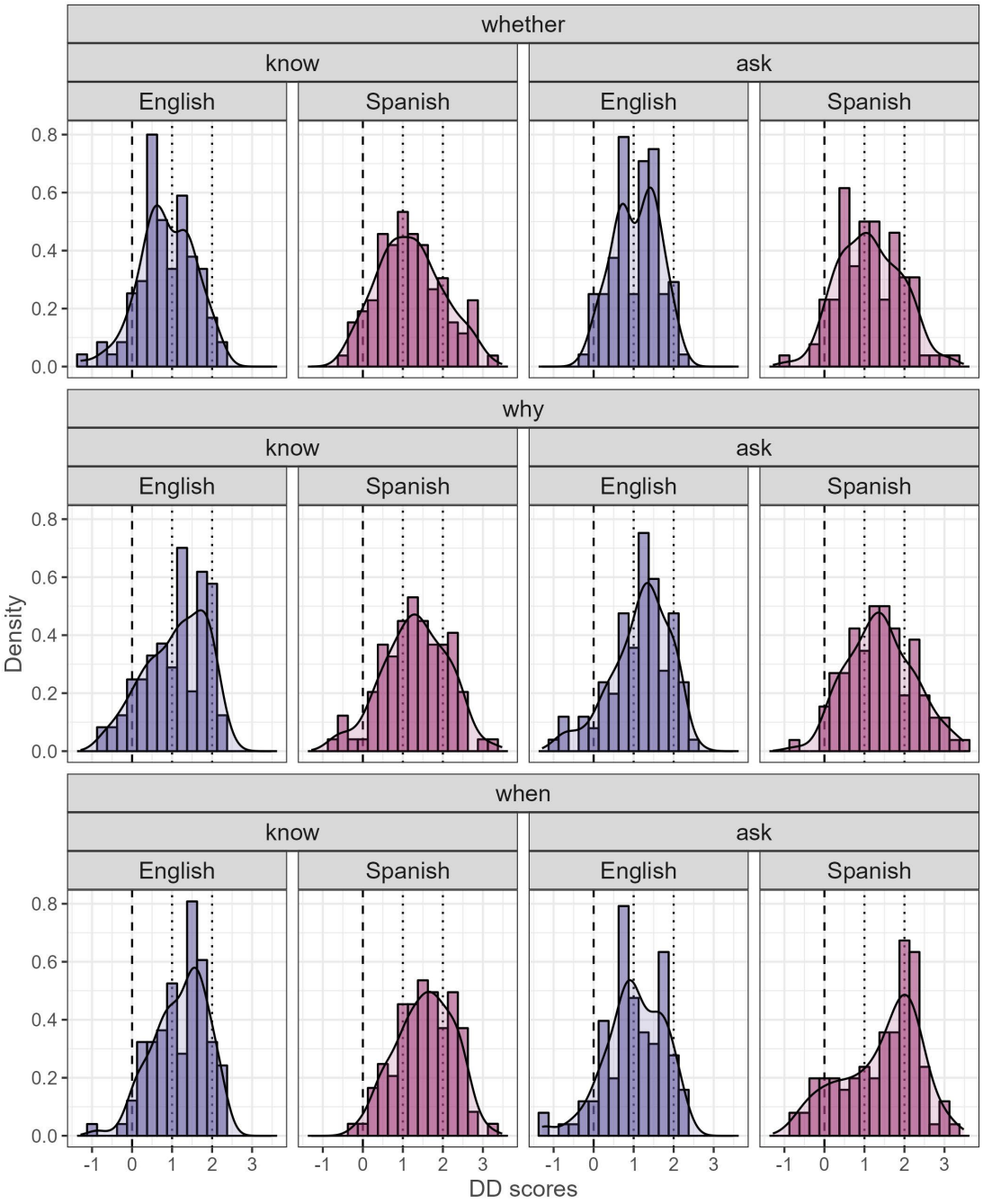

Figure 3: Distribution of by-participant DD scores in the English and Spanish experiments. The dashed vertical line marks the limit between DD scores higher than 0 (indicative of an island effect) and DD scores lower than 0 (indicative of no island effect).

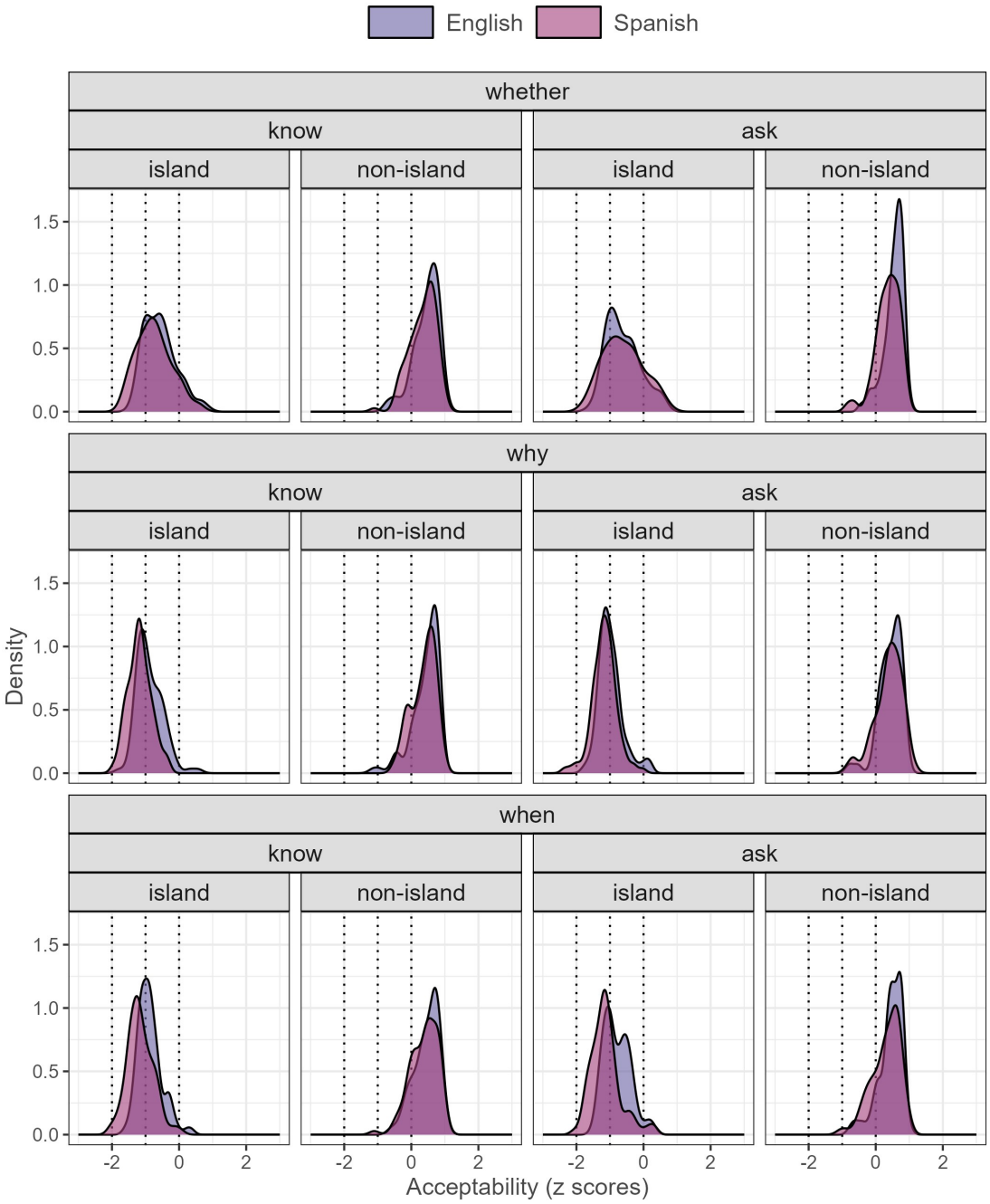

Figure 4: Distribution of by-participant z-score means in the island/embedded and non-island/embedded conditions of the English and Spanish experiments.

We tested for bimodality statistically using the multimode package (Ameijeiras-Alonso et al. 2021) in R (R Core Team 2023). We chose three tests that instantiate the most common approaches to identifying multimodality: the dip test (Hartigan & Hartigan 1985), the excess mass test (Müller & Sawitzki 1991; Cheng & Hall 1998; Ameijeiras-Alonso et al. 2019), and the bandwidth test (Silverman 1981; Hall & York 2001). What we looked for is a statistically significant effect of multimodality in Spanish but not in English in either the DD scores or the individual condition z-scores. The full list of *p*-values for each of these tests is shown in Table 5. Crucially, there are no statistically significant effects for any of the Spanish islands under any of the tests, corroborating the visual inspection of Figure 3 and Figure 4, and suggesting that there is no evidence of two or more populations of speakers of Spanish in our studies. In English, there are significant effects in the *ask whether* DD scores and in the *know whether* and *ask when* island/embedded condition, both under the excess mass test (a significant effect is also observed in the English *ask when* non-island/embedded condition). If anything, these results suggest that English is more variable than Spanish, contrary to the hypothesis we set out to explore.

	Dip test		Excess mass test		Bandwidth test	
	English	Spanish	English	Spanish	English	Spanish
DD scores						
know whether	.256	.836	.060	.454	.308	.226
ask whether	.138	.410	**.026**	.120	.066	.600
know why	.370	.900	.118	.632	.516	.368
ask why	.748	.460	.220	.080	.506	.438
know when	.986	.994	.862	.908	.310	.874
ask when	.714	.998	.314	.966	.312	.774
know whether	.256	.836	.060	.454	.308	.226
island/embedded						
know whether	.302	.944	**.036**	.592	.124	.558
ask whether	.556	.810	.164	.382	.278	.540
know why	1	1	.994	.974	.238	.508
ask why	.998	.962	.936	.632	.214	.884
know when	.958	.786	.626	.318	.292	.400
ask when	.086	.864	**.008**	.398	.076	.316
non-island/embedded						
know whether	.622	.932	.150	.596	.358	.128
ask whether	.998	.506	.956	.142	.748	.116
know why	.986	.672	.832	.256	.344	.128
ask why	.670	.852	.186	.348	.088	.190
know when	.962	.438	.638	.126	.846	.138
ask when	.260	.976	.020	.746	.406	.322

Table 5: *P*-values obtained in three multimodality tests (the dip test, the excess mass test and the bandwidth test) run on the DD scores, the island/embedded conditions and the non-island/embedded conditions (which are shown as a control).

Within-participant variation is displayed in Figure 5 and Figure 6. Figure 5 plots each of the 4 ratings of the island/embedded condition for each participant: each horizontal line represents a participant; their individual ratings are represented by the colored dots and their mean rating is represented by the black circles. We categorized the four ratings of the island/embedded condition given by each participant as either below or above their mid-point of the scale (0). Then, we counted how many participants rated all four tokens below 0, how many rated 3 out of 4 below 0, etc. The end result are counts for 5 categories of speakers for each of the two languages, which we then converted to proportions of the total sample for each experiment. These proportions are shown as a stacked bar plot in Figure 6. Visual inspection of the proportions in Figure 6 suggests that the two languages show similar amounts of within-participant variability.

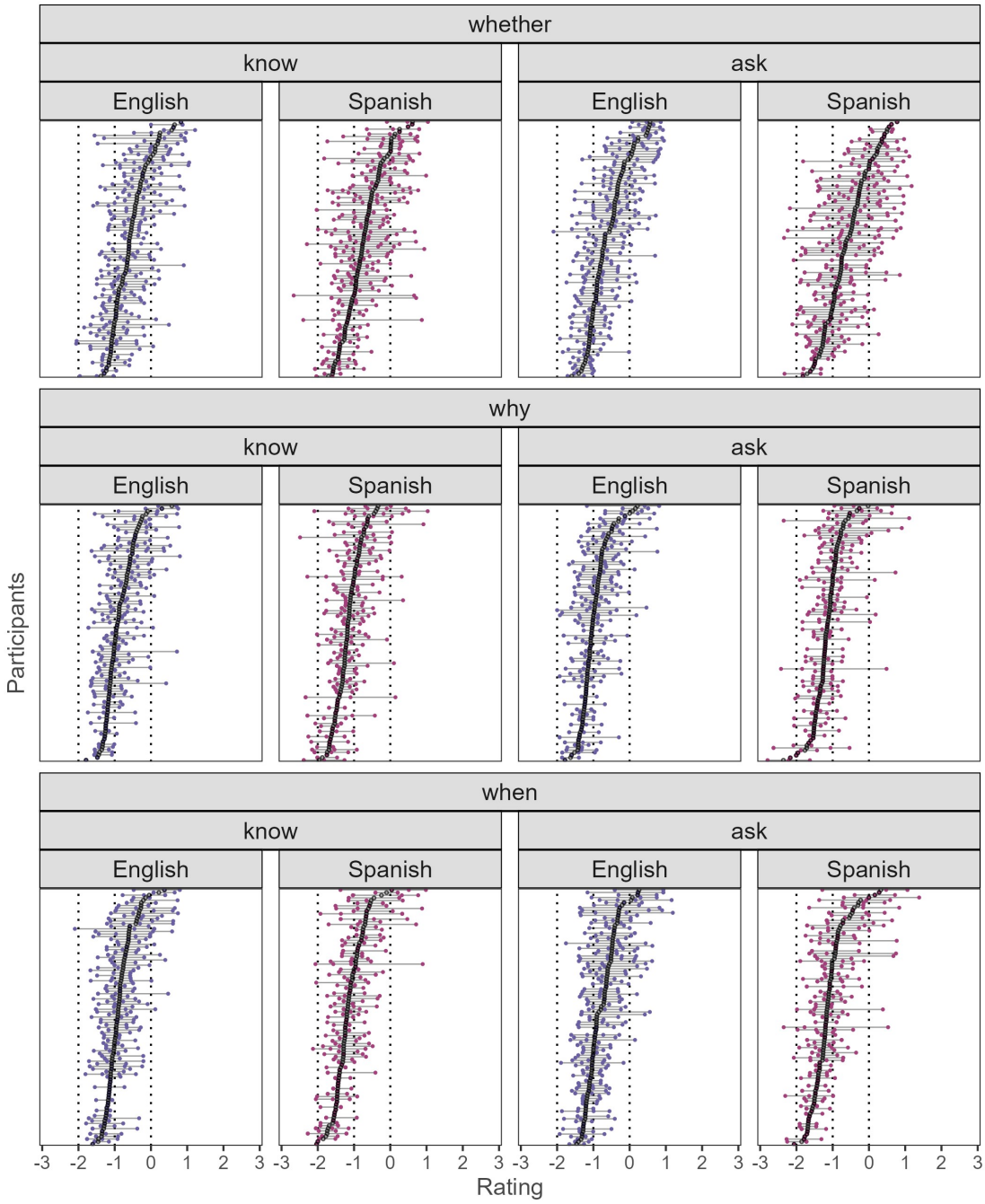

Figure 5: Distribution of by-participant z-scores in the island/embedded conditions of all the experiments. Each horizontal line represents a participant; the dots on each line indicate the location of each of the participant's observations on the z-score scale, and the black circles show the participants' mean (for each experiment, participants are ordered by their mean z-score).

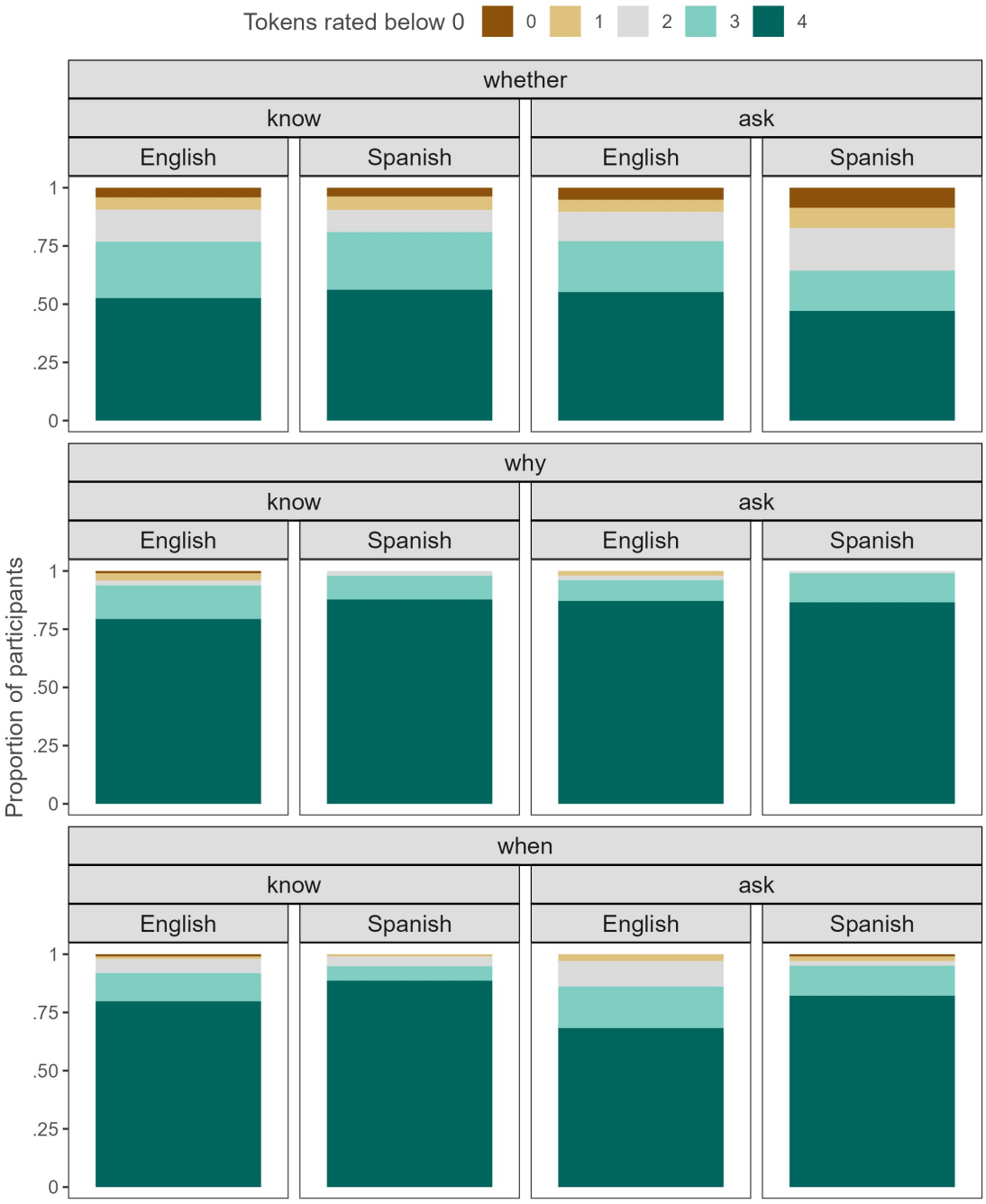

Figure 6: Proportion of participants that rated 0, 1, 2, 3 or 4 out of 4 tokens below their midpoint of the z-score scale (0) in each of the experiments.

We tested whether the counts of the 5 categories differed statistically between the two languages by using Fisher's exact test (on a 2×5 contingency table crossing language and the 5 count categories). We found that only *ask when* yielded a significant effect (p = .028), while the other five island types were non-significant (*know when*: p = .415; *ask whether*: p = .429; *know whether*: p = .922; *ask why*: p = .477; *know why*: p = .282). As we can see in Figure 6, the significant difference for *ask when* appears to be driven by English showing more within-participant variability (fewer participants with 4/4 ratings below 0) than Spanish. This runs contrary to the hypothesis that perhaps Spanish shows more within-participant variability.

Taken together, the results of these two analyses suggest that there is no more between- or within-participant variation in Spanish than in English, at least for the participants in our experiments.

5 Discussion

5.1 Wh-island effects in Spanish and English

In a series of 12 acceptability judgment experiments, we examined three types of *wh*-island effects (*whether*, *why* and *when* island effects) under two embedding verbs (*know* and *ask*) in English and Spanish translation-equivalent sentences with *wh*-dependencies. The first question we sought to address was whether Spanish *wh*-extraction showed island effects in these contexts. The answer to this question appears to be yes – we find clear evidence of island effects for the participants that volunteered for our study (i.e., users of Prolific (www.prolific.co) in 2023).

The second question we sought to address was whether there is a gradient difference in effect size between *wh*-islands in English and Spanish. One possibility we considered was that the locus of cross-linguistic variation lies in effect size rather than the categorical presence/absence of island effects, perhaps with Spanish showing smaller island effects than English. What we found is that there is a difference in effect size for 4 or 5 out of the 6 island types that we tested, but it goes in an unexpected direction – Spanish island effects tend to be slightly larger than English island effects (with translation-equivalent materials). To our knowledge, there are no theories of differences in effect sizes for island effects that we can use to interpret this effect. We do note that the experimental literature has cataloged a number of results showing minor variation in the island effect sizes across languages (Sprouse & Villata 2021 for a review), which may be a productive avenue for future theorizing.

The final question that we sought to address was whether there is increased individual variation in Spanish as compared to English (both between- and within-participants), as this could be a potential explanation for some of the variability in the observations reported in the literature on Spanish. We find no evidence of greater between- or within-participant variability in Spanish (and, in fact, the only island that showed increased variability was in English). This provides additional support for the similarity between the two languages. The question whether there could be more variability in Spanish than English was reasonable given the contrast between Torrego's (1984) observations and recent experimental findings, and given that experimental studies have found increased between and within-participant variability, both in other languages (e.g., Norwegian; see Kush et al. 2018; 2019, Kobzeva et al. 2022) and in Spanish *when wh*-islands with "d-linked" fillers (Pañeda & Kush 2022). However, the Spanish speakers in our study do not show more variation than the set of English

speakers. This finding helps to underscore that the island effects that we observed are likely a robust part of Spanish for these speakers. On that note, it seems valuable for future work to continue to probe individual variation because it can potentially reveal more information on factors that govern acceptability (cognitive, dialectal, etc.) or help to establish that the effect is indeed robust.

5.2 Theoretical implications

Our findings lead us to conclude that *wh*-extraction gives rise to island effects in both English and Spanish and that these effects are not smaller in Spanish. This suggests that the two languages share at least one constraint that causes or contributes to the phenomenon. Our study was not designed to determine the source of *wh*-island effects, so our results are compatible with different theories. In what follows, we discuss how they would fit in syntactic, semantic, processing, and information-structure accounts.

One of the most prominent syntactic accounts of *wh*-island effects was based on the principle of Subjacency, which bans movement steps that cross more than one bounding node (Chomsky 1973; 1977). It was proposed that in English Inflectional Phrase (IP) is a bounding node, which causes extraction from *wh*-islands to violate Subjacency (Chomsky 1977): the reason is that there are two IPs between the base position of the extractee and its landing position, and they cannot be crossed in separate movement steps given that the intermediate Spec, Complementizer Phrase (CP) position that could provide an "escape hatch" is filled with the *wh*-word. In contrast, Spanish and Italian were argued to have CP rather than IP as a bounding node, which enabled extraction from *wh*-islands to comply with Subjacency, as there is only one CP node to cross between the base position of the extractee and the landing position (Rizzi 1982; Torrego 1984). Our finding that both English and Spanish show *wh*-island effects might suggest that IP counts as a bounding node for both languages.[2]

Another prominent syntactic account attributes *wh*-island effects to a violation of Relativized Minimality (RM) (Rizzi 1990). RM holds that an extractee can only establish a dependency with its base position if there is no intervener that could engage in the same dependency, where intervener is defined as a constituent that shares relevant morphosyntactic features with the extractee (such as a *wh*-feature), c-commands the base position from the same type of position as the landing position (such as spec, CP), but crucially does not c-command the landing position (see also Friedmann et al. 2009; Belletti et al. 2012; Rizzi 2013; Atkinson et al. 2016; Villata et al. 2016). The *wh*-word introducing *wh*-islands meets the requirements of an intervener, causing extraction from *wh*-clauses to violate RM. In this framework, cross-linguistic variation in *wh*-island effects could be accounted for (i) if the intervener shared a movement feature with the extractee in one language and not in the other, or (ii) if the structural position of the intervener was the same type as the landing position in one language but not the other. However, our findings do not support any such differences, and instead could be taken to suggest that both the English and the Spanish *wh*-phrases we used have a similar featural composition that causes them to induce *wh*-island effects.

---

[2] Similar considerations apply to the *Barriers* (Chomsky 1986) formulation of Subjacency. In this framework, it was proposed that CP was a barrier for movement in both English and Italian/Spanish, but that in English tensed IP was also a barrier, causing extraction from *wh*-islands to violate Subjacency. Our findings could suggest that tensed IP counts as an extra barrier in both languages.

Wh-island effects have also been explained from a semantic perspective. One relevant proposal (Szabolsci & Zwarts 1993) attributes some island effects to difficulties associated with interpreting the extractee in the domain of 'scopal interveners' such as *wh*-phrases at the edge of embedded questions. Under this account, extractees that range over *individuals* can be interpreted in the domain of *wh*-interveners because all Boolean operations for individuals are defined in those domains. By contrast, extractees that range over ordered sets (e.g., properties) or are non-referential cannot be interpreted in the domain of a *wh*-intervener because the compositional operations required for computing the relevant interpretations are not defined in that domain (See also Abrusán 2014 for a similar proposal). Under this account, our findings would suggest that speakers of both languages tend to adopt a *property* interpretation of the extractee instead of an *individual* interpretation.

Wh-island effects may alternatively be viewed as resulting from processing difficulty. Under a maintenance theory of working memory, this difficulty may arise because processing the dependency and the embedded island structure exceeds the available working memory capacity, causing working memory overload (Kluender & Kutas 1993; Hofmeister & Sag 2010). Alternatively, under a cue-based theory of working memory, the difficulty may arise because the intervener is similar to the extractee in its featural composition and therefore interferes with the resolution of the dependency. There are two ways in which the intervener may interfere: (i) by hindering the encoding of the extractee in working memory when it is encountered at the landing position during left-to-right processing, or (ii) by hindering the retrieval of the extractee from working memory to resolve the dependency at the base position (Atkinson et al. 2016; Villata et al. 2016; Keshev & Meltzer-Asscher 2019). It has been proposed that Spanish could show no or smaller *wh*-island effects than English because it has a richer morphology that provides more informative encoding and/or retrieval cues, facilitating processing (Ortega-Santos 2011). However, since we did not find evidence that Spanish has smaller *wh*-island effects than English, our results do not support this possibility: there is no evidence of a difference in either the encoding/retrieval features or the processing cost of the embedded *wh*-island structure.

Finally, under information-structure-based theories, *wh*-island effects arise when the embedded *wh*-clauses are backgrounded, which creates a clash with the focused extractee (Erteschik-Shir 1973; Goldberg 2005; 2013; Abeillé et al. 2020; Cuneo & Goldberg 2023). From this perspective, cross-linguistic variation could arise if *wh*-clauses are backgrounded in some languages and (more) focused in others (see also Erteschik-Shir 1973 for variation in *wh*-islands between English and Danish). Within this framework, our results would suggest that *wh*-islands are backgrounded constituents in both English and Spanish.

Though our studies were not designed to uniquely determine the source of *wh*-island effects, our results do suggest that there is no evidence that *wh*-island effects with *wh*-extraction are less present in Spanish than English. One theoretical consequence of this is that the English/Spanish contrast can no longer be seen as providing unequivocal empirical support for a parameters-based theory of cross-linguistic variation (Rizzi 1982; Torrego 1984), despite the prominent role that the purported contrast played in the theory's development. It should be noted, though, that the results are still in principle compatible with parameterized syntactic variation if the island effects we observe stem from extra-syntactic sources, which could conceal potential syntactic differences. Since

our study was not designed to test the source of wh-island effects explicitly, future work could explore configurations where the accounts make differing predictions.

Overall, our results align with previous findings that the full picture of cross-linguistic variation in *wh*-island effects may be more complicated than previously thought and that *wh*-island effects are more pervasive than previously thought (e.g., see Almeida 2014; Sprouse et al. 2016; Kush et al. 2018). More languages will need to be studied to obtain a more complete picture of the space of possible theories of cross-linguistic variation.

6 Conclusion

In a series of 12 acceptability judgment experiments, we examined *wh*-extraction from three types of embedded *wh*-questions (*whether*, *why* and *when* island effects) under two embedding verbs (*know* and *ask*) in both Spanish and English sentences. Our goal was to explore the original observation by Torrego (1984) that Spanish does not show *wh*-island effects in sentences with object *wh*-extraction from embedded *wh*-questions introduced by non-arguments. We found: (i) *wh*-island effects for both languages for all 6 island types tested, (ii) larger island effects for Spanish compared to English for most of the island types tested, and (iii) no evidence of additional between- or within-participant variation for Spanish compared to English. These results run contrary to both the original, binary version of the cross-linguistic claim and a plausible gradient variant. Our findings replicate previous experimental work showing that there are island effects in both languages (Sprouse 2007; Sprouse et al. 2011; 2012; 2016; Almeida 2014; Michel 2014; Aldosari 2015; López-Sancio 2015; Ortega-Santos et al. 2018; Pañeda et al. 2020; Pham et al. 2020; Rodríguez & Goodall 2020; Stigliano & Xiang 2021; Pañeda & Kush 2022), and also extend it to a wider range of relevant island types, using relatively large samples of participants, and translation-equivalent materials. Our results suggest that the European Spanish spoken by the participants that volunteered for our study is clearly a *wh*-island language, and therefore that the use of European Spanish as evidence for theories that encode cross-linguistic variation in *wh*-island effects may need to be reconsidered.

Abbreviations

SG: singular; PST: past; PL: plural; COND: conditional; F: feminine; M: masculine; DD: differences-in-differences; BF: Bayes Factor; SE: standard error; d-linked: discourse-linked; IP: inflectional phrase; CP: complementizer phrase; RM: Relativized Minimality.

Supplementary files

S1: Materials

English experimental items

(1)	
non-island/matrix	Who thought that they would modify the recipe?
non-island/embedded	What did the cook think that they would modify?
island/matrix	Who asked/wanted to know whether/why/when they would modify the recipe?
island/embedded	What did the cook ask/want to know whether/why/when they would modify?
(2)	
non-island/matrix	Who thought that you would refurbish the room?
non-island/embedded	What did the decorator think that you would refurbish?
island/matrix	Who asked/wanted to know whether/why/when you would refurbish the room?
island/embedded	What did the decorator ask/want to know whether/why/when you would refurbish?
(3)	
non-island/matrix	Who thought that they would reject the proposal?
non-island/embedded	What did the politician think that they would reject?
island/matrix	Who asked/wanted to know whether/why/when they would reject the proposal?
island/embedded	What did the politician ask/want to know whether/why/when they would reject?
(4)	
non-island/matrix	Who thought that they would demolish the building?
non-island/embedded	What did the architect think that they would demolish?
island/matrix	Who asked/wanted to know whether/why/when they would demolish the building?
island/embedded	What did the architect ask/want to know whether/why/when they would demolish?

(5)

non-island/matrix	Who thought that you would rewrite the text?
non-island/embedded	What did the journalist think that you would rewrite?
island/matrix	Who asked/wanted to know whether/why/when you would rewrite the text?
island/embedded	What did the journalist ask/want to know whether/why/when you would rewrite?

(6)

non-island/matrix	Who thought that they would cancel the show?
non-island/embedded	What did the actor think that they would cancel?
island/matrix	Who asked/wanted to know whether/why/when they would cancel the show?
island/embedded	What did the actor ask/want to know whether/why/when they would cancel?

(7)

non-island/matrix	Who thought that they would sanitize the machine?
non-island/embedded	What did the scientist think that they would sanitize?
island/matrix	Who asked/wanted to know whether/why/when they would sanitize the machine?
island/embedded	What did the scientist ask/want to know whether/why/when they would sanitize?

(8)

non-island/matrix	Who thought that they would inspect the suitcase?
non-island/embedded	What did the passenger think that they would inspect?
island/matrix	Who asked/wanted to know whether/why/when they would inspect the suitcase?
island/embedded	What did the passenger ask/want to know whether/why/when they would inspect?

(9)

non-island/matrix	Who thought that you would support the reform?
non-island/embedded	What did the congressman think that you would support?
island/matrix	Who asked/wanted to know whether/why/when you would support the reform?
island/embedded	What did the congressman ask/want to know whether/why/when you would support?

(10)

non-island/matrix	Who thought that you would fund the project?
non-island/embedded	What did the businessman think that you would fund?
island/matrix	Who asked/wanted to know whether/why/when you would fund the project?
island/embedded	What did the businessman ask/want to know whether/why/when you would fund?

(11)

non-island/matrix	Who thought that you would challenge the ruling?
non-island/embedded	What did the lawyer think that you would challenge?
island/matrix	Who asked/wanted to know whether/why/when you would challenge the ruling?
island/embedded	What did the lawyer ask/want to know whether/why/when you would challenge?

(12)

non-island/matrix	Who thought that they would bandage the wound?
non-island/embedded	What did the nurse think that they would bandage?
island/matrix	Who asked/wanted to know whether/why/when they would bandage the wound?
island/embedded	What did the nurse ask/want to know whether/why/when they would bandage?

(13)

non-island/matrix	Who thought that you would submit the application?
non-island/embedded	What did the secretary think that you would submit?
island/matrix	Who asked/wanted to know whether/why/when you would submit the application?
island/embedded	What did the secretary ask/want to know whether/why/when you would submit?

(14)

non-island/matrix	Who thought that they would trim the bush?
non-island/embedded	What did the gardener think that they would trim?
island/matrix	Who asked/wanted to know whether/why/when they would trim the bush?
island/embedded	What did the gardener ask/want to know whether/why/when they would trim?

(15)

non-island/matrix	Who thought that you would prepare the report?
non-island/embedded	What did the clerk think that you would prepare?
island/matrix	Who asked/wanted to know whether/why/when you would prepare the report?
island/embedded	What did the clerk ask/want to know whether/why/when you would prepare?

(16)

non-island/matrix	Who thought that you would tune the instrument?
non-island/embedded	What did the musician think that you would tune?
island/matrix	Who asked/wanted to know whether/why/when you would tune the instrument?
island/embedded	What did the musician ask/want to know whether/why/when you would tune?

English fillers

(17) Mike prefers tennis because Jon baseball.
(18) Jenny cleaned her sister the table.
(19) Forget to walk, Susan did the dog.
(20) The was painted vase carelessly.
(21) There had all hung over the fireplace the portraits by Picasso.
(22) Lily will dance who the king chooses.
(23) Sue asked what who bought.
(24) The employees were overworked and the executives by the managers.
(25) The specimen thawed to study it more closely.
(26) With that announcement were many citizens denied the opportunity to protest.
(27) The flowers plant seasonally to attract hummingbirds.
(28) It is not wise one to skip class.
(29) Anyone better come home.
(30) There is likely a river to run down the mountain.
(31) Richard may have been hiding, but Blake may have done so too.
(32) Tom argued with as stubborn a man as his brother.
(33) They all have helped and they have done all so generously.
(34) Lloyd Webber musicals are easy to condemn without even watching.
(35) The ball perfectly rolled down the hill.
(36) John put more books than Bill did on the table.
(37) She invited Briana and ignored her.
(38) It seemed at that time Robert had confessed.
(39) There are firemen injured.
(40) Someone better sing the national anthem.
(41) Fred baked the kids coconut cookies.
(42) One mechanic managed to repair every car.
(43) Laura is more excited than nervous.
(44) I hate eating sushi.
(45) Marcie was fired by her manager.
(46) Martin tried to solve the math problem.
(47) There might mice seem to be in the cupboard.
(48) It seems to me that Robert can't be trusted.

English practice items

(49) Promise to wash, Neal did the car.
(50) Ben is hopeful for everyone you do to attend.
(51) They consider a teacher of Chris geeky.
(52) The brother and sister that were playing all the time had to be sent to bed.
(53) All the men seem to have all eaten supper.
(54) The children were cared for by the adults and the teenagers.

(55) She was the winner.

Spanish experimental items

[For the translations, see the corresponding English items.]

(1)	
non-island/matrix	¿Quién pensaba que modificarían la receta?
non-island/embedded	¿Qué pensaba el cocinero que modificarían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué modificarían la receta?
island/embedded	¿Qué preguntó/quería saber el cocinero si modificarían?
(2)	
non-island/matrix	¿Quién pensaba que modernizarías la habitación?
non-island/embedded	¿Qué pensaba la decoradora que modernizarías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué modernizarías la habitación?
island/embedded	¿Qué preguntó/quería saber la decoradora si/cuándo/por qué modernizarías?
(3)	
non-island/matrix	¿Quién pensaba que rechazarían la propuesta?
non-island/embedded	¿Qué pensaba el político que rechazarían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué rechazarían la propuesta?
island/embedded	¿Qué preguntó/quería saber el político si/cuándo/por qué rechazarían?
(4)	
non-island/matrix	¿Quién pensaba que demolerían el edificio?
non-island/embedded	¿Qué pensaba el arquitecto que demolerían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué demolerían el edificio?
island/embedded	¿Qué preguntó/quería saber el arquitecto si/cuándo/por qué demolerían?
(5)	
non-island/matrix	¿Quién pensaba que reescribirías el texto?
non-island/embedded	¿Qué pensaba la periodista que reescribirías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué reescribirías el texto?
island/embedded	¿Qué preguntó/quería saber la periodista si/cuándo/por qué reescribirías?

(6)

non-island/matrix	¿Quién pensaba que anularían el espectáculo?
non-island/embedded	¿Qué pensaba el actor que anularían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué anularían el espectáculo?
island/embedded	¿Qué preguntó/quería saber el actor si/cuándo/por qué anularían?

(7)

non-island/matrix	¿Quién pensaba que desinfectarían la máquina?
non-island/embedded	¿Qué pensaba el científico que desinfectarían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué desinfectarían la máquina?
island/embedded	¿Qué preguntó/quería saber el científico si/cuándo/por qué desinfectarían?

(8)

non-island/matrix	¿Quién pensaba que inspeccionarían la maleta?
non-island/embedded	¿Qué pensaba el pasajero que inspeccionarían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué inspeccionarían la maleta?
island/embedded	¿Qué preguntó/quería saber el pasajero si/cuándo/por qué inspeccionarían?

(9)

non-island/matrix	¿Quién pensaba que apoyarías la reforma?
non-island/embedded	¿Qué pensaba el diputado que apoyarías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué apoyarías la reforma?
island/embedded	¿Qué preguntó/quería saber el diputado si/cuándo/por qué apoyarías?

(10)

non-island/matrix	¿Quién pensaba que financiarías el proyecto?
non-island/embedded	¿Qué pensaba el empresario que financiarías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué financiarías el proyecto?
island/embedded	¿Qué preguntó/quería saber el empresario si/cuándo/por qué financiarías?

(11)

non-island/matrix	¿Quién pensaba que cuestionarías la sentencia?
non-island/embedded	¿Qué pensaba la abogada que cuestionarías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué cuestionarías la sentencia?
island/embedded	¿Qué preguntó/quería saber la abogada si/cuándo/por qué cuestionarías?

(12)

non-island/matrix	¿Quién pensaba que vendarían la herida?
non-island/embedded	¿Qué pensaba la enfermera que vendarían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué vendarían la herida?
island/embedded	¿Qué preguntó/quería saber la enfermera si/cuándo/por qué vendarían?

(13)

non-island/matrix	¿Quién pensaba que enviarías la solicitud?
non-island/embedded	¿Qué pensaba la secretaria que enviarías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué enviarías la solicitud?
island/embedded	¿Qué preguntó/quería saber la secretaria si/cuándo/por qué enviarías?

(14)

non-island/matrix	¿Quién pensaba que podarían el arbusto?
non-island/embedded	¿Qué pensaba el jardinero que podarían?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué podarían el arbusto?
island/embedded	¿Qué preguntó/quería saber el jardinero si/cuándo/por qué podarían?

(15)

non-island/matrix	¿Quién pensaba que prepararías el informe?
non-island/embedded	¿Qué pensaba el administrativo que prepararías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué prepararías el informe?
island/embedded	¿Qué preguntó/quería saber el administrativo si/cuándo/por qué prepararías?

(16)

non-island/matrix	¿Quién pensaba que afinarías el instrumento?
non-island/embedded	¿Qué pensaba el músico que afinarías?
island/matrix	¿Quién preguntó/quería saber si/cuándo/por qué afinarías el instrumento?
island/embedded	¿Qué preguntó/quería saber el músico si/cuándo/por qué afinarías?

Spanish fillers

(17) ¿Y tú no sabes quién llegó cómo?
'And you don't know who arrived how?'

(18) A Mara, Pepe conoció al muchacho que la quiere.
'Mara, Pepe met the guy who loves her.'

(19) ¿Quién me dijiste que a tu madre que la va a llamar?
'Who did you tell me that your mother that she is going to call her?'

(20) Resolvimos el problema poco complicado más interesante del libro.
'We solved the little complicated problem more interesting in the book.'
(21) Te pido que a tu padre en este momento esa mentira que no se la digas.
'I ask you that your dad at this point that lie that you don't tell.'
(22) La monja, dicen que no le van a dar nada a ella.
'The nun, they say that they are not going to give her anything.'
(23) ¿Iván le pidió qué el otro día a quién?
'Iván asked for what the other day to whom?'
(24) ¡Oye, tú, bastante guapa!
'Hey, you, rather pretty!'
(25) He leído el libro un poco interesante de Rómulo Gallegos.
'I read the slightly interesting book by Rómulo Gallegos.'
(26) ¿Y tú le diste a quién la guitarra?
'And you gave the guitar to whom?'
(27) Pedro es como tú de gordo para su estatura.
'Pedro is similarly fat as you for his height.'
(28) ¡Lo mínimamente preocupado que está Pedro!
'How minimally worried is Pedro!'
(29) María y Ana son poco altas.
'María and Ana are little tall.'
(30) Gritaron que que se cancela la fiesta si llueve.
'They shouted that that the party will be cancelled if it rains.'
(31) Baila con la chica altísima.
'(S)he dances with the very tall girl.'
(32) ¿A cuántos de ellos querían nombrarlos sin conocer?
'How many of them did they want to name without knowing?'
(33) Juan es como yo de feo para ser modelo.
'Juan is similarly ugly as I to be a model.'
(34) Pedro busca a un secretario que habla inglés.
'Pedro is looking for a secretary that speaks English.'
(35) Cantar Pavarotti en el Liceo fue maravilloso.
'Singing Pavarotti at the Liceo was marvellous.'
(36) Hazlo tú.
'Do it yourself.'
(37) Este libro es muy poco interesante.
'That book is very little interesting.'
(38) Juan dejó de tomar el café.
'Juan stopped drinking the coffee.'
(39) A una candidata, el jefe la descartó sin entrevistar.
'A candidate, the boss ruled out without interviewing.'
(40) Nos han pillado a todos nosotros.
'They caught us all.'
(41) Vimos algunos libros.
'We saw some books.'
(42) Quiero a demasiadas mujeres.
'I love too many women.'

(43) Ojalá que venga.
'Hopefully (s)he will come.'
(44) Mi abuelo fuma puros.
'My grandpa smokes cigars.'
(45) Esta mesa es más alta que la mía.
'This table is taller than mine.'
(46) Cayeron hojas.
'Leaves fell.'
(47) María llegó con un señor un montón.
'María arrived with a gentleman a lot.'
(48) Pedro tocó la mesa.
'Pedro touched the table.'

Spanish practice items

(49) Esto prueba cantar tú la Traviata muy bien.
'This proves you to sing the Traviata very good.'
(50) ¡Al perro, que lo bañen, y al gato, que también!
'The dog, may they bath it, and the cat that too!
(51) Dice que si llueve, entonces, venga.
'(S)he says that if it rains, then, (s)he should come.
(52) Como lo que hay en la nevera.
'I eat what is in the fridge.'
(53) Me susurró que lo hizo.
'(S)he whispered to me that (s)he did it.'
(54) Los científicos estudian el asunto.
'The scientist study the matter.'
(55) Juana leyó durante una hora.
'Juana read for an hour.'

S2: Additional results

	Estimate	SE	t	p	BF
Know whether					
Intercept	0.209	0.026	8.178	< .001	
Structure	−0.266	0.013	−20.076	< .001	
Position	−0.342	0.017	−19.689	< .001	
Language	−0.089	0.021	−4.186	< .001	
Structure × Position	−0.256	0.009	−27.424	< .001	
Structure × Language	0.024	0.013	1.833	.068	
Position × Language	0.018	0.014	1.289	.199	
Structure × Position × Language	−0.040	0.009	−4.259	< .001	62
Ask whether					
Intercept	0.269	0.028	9.671	< .001	
Structure	−0.256	0.013	−19.666	< .001	
Position	−0.348	0.016	−22.301	< .001	
Language	−0.062	0.020	−3.049	.005	
Structure × Position	−0.275	0.009	−30.505	< .001	
Structure × Language	0.056	0.013	4.340	< .001	
Position × Language	0.017	0.015	1.178	.240	
Structure × Position × Language	−0.009	0.009	−1.020	.308	0.12

Table 6: Results of the Structure × Position × Language linear mixed models for *know whether* and *ask whether* islands, and Bayes Factors (BF) for the Structure × Position × Language interactions. The factors followed an effects coding scheme: Structure (non-island: −1, island: 1), Position (matrix: −1, embedded: 1), Language (English: −1, Spanish: 1). The models included random intercepts and structure and position slopes for participant and random intercepts and position and language slopes for item. The interactions were not included in the slopes to ensure convergence.

	Estimate	SE	t	p	BF
Know why					
Intercept	0.072	0.022	3.206	.004	
Structure	−0.414	0.014	−30.084	< .001	
Position	−0.387	0.017	−22.348	< .001	
Language	−0.141	0.016	−8.562	< .001	
Structure × Position	−0.299	0.009	−33.421	< .001	
Structure × Language	−0.031	0.014	−2.275	.024	
Position × Language	0.051	0.014	3.727	< .001	
Structure × Position × Language	−0.027	0.009	−3.018	.003	3.12
Ask why					
Intercept	0.076	0.022	3.429	.003	
Structure	−0.422	0.014	−30.204	< .001	
Position	−0.411	0.017	−23.842	< .001	
Language	−0.113	0.016	−6.939	< .001	
Structure × Position	−0.320	0.009	−35.585	< .001	
Structure × Language	−0.003	0.012	−0.261	.794	
Position × Language	0.054	0.013	4.247	< .001	
Structure × Position × Language	−0.030	0.009	−3.325	.001	7.31

Table 7: Results of the Structure × Position × Language linear mixed models for *know why* and *ask why* islands, and Bayes Factors (BF) for the Structure × Position × Language interactions. The factors followed an effects coding scheme: Structure (non-island: −1, island: 1), Position (matrix: −1, embedded: 1), Language (English: −1, Spanish: 1). Both models included random intercepts and structure and position slopes for participant and random intercepts and position and language slopes for item. The *know why* model additionally included a structure slope. The interactions were not included in the slopes to ensure convergence.

	Estimate	SE	t	p	BF
Know when					
Intercept	0.133	0.025	5.262	< .001	
Structure	−0.370	0.013	−27.714	< .001	
Position	−0.433	0.016	−27.627	< .001	
Language	−0.117	0.017	−7.019	< .001	
Structure × Position	−0.344	0.009	−40.219	< .001	
Structure × Language	−0.007	0.013	−0.520	.604	
Position × Language	0.028	0.012	2.275	.024	
Structure × Position × Language	−0.044	0.009	−5.174	< .001	> 100
Ask when					
Intercept	0.135	0.025	5.458	< .001	
Structure	−0.359	0.019	−18.986	< .001	
Position	−0.429	0.018	−23.424	< .001	
Language	−0.103	0.014	−7.138	< .001	
Structure × Position	−0.305	0.009	−33.219	< .001	
Structure × Language	0.000	0.013	−0.015	.988	
Position × Language	−0.015	0.014	−1.046	.297	
Structure × Position × Language	−0.052	0.009	−5.693	< .001	> 100

Table 8: Results of the Structure × Position × Language linear mixed models for *know when* and *ask when* islands, and Bayes Factors (BF) for the Structure × Position × Language interactions. The factors followed an effects coding scheme: Structure (non-island: −1, island: 1), Position (matrix: −1, embedded: 1), Language (English: −1, Spanish: 1). Both models included random intercepts and structure and position slopes for participant and random intercepts and structure, position and language slopes for item. The interactions were not included in the slopes to ensure convergence.

Data availability

The data and data analysis script are available at https://osf.io/xztvy/?view_only=274bd9e9dca44c498caf5c743bcec656

Ethics and consent

This worked was approved by the NYUAD Institutional Review Board (approval number: HRPP-2021-117). All procedures were in accordance with the Declaration of Helsinki.

Competing interests

The authors have no competing interests to declare.

References

Abeillé, Anne & Hemforth, Barbara & Winckel, Elodie & Gibson, Edward. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204. 104293. DOI: https://doi.org/10.1016/j.cognition.2020.104293

Abrusán, Márta. 2014. *Weak island semantics*. New York, NY: Oxford University Press.

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.

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

Ameijeiras-Alonso, José & Crujeiras, Rosa M. & Rodríguez-Casal, Alberto. 2019. Mode testing, critical bandwidth and excess mass. *Test* 28. 900–919. DOI: https://doi.org/10.1007/s11749-018-0611-5

Ameijeiras-Alonso, José, & Crujeiras, Rosa M. & Rodríguez-Casal, Alberto. 2021. multimode: An R Package for Mode Assessment. *Journal of Statistical Software* 97(9). 1–32. DOI: https://doi.org/10.18637/jss.v097.i09

Atkinson, Emily & Apple, Aaron & Rawlins, Kyle & Omaki, Akira. 2016. Similarity of *wh*-phrases and acceptability variation in *wh*-islands. *Frontiers in Psychology* 6. 2048. DOI: https://doi.org/10.3389/fpsyg.2015.02048

Belletti, Adriana & Friedmann, Naama & Brunato, Dominique & Rizzi, Luigi. 2012. Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua* 122(10). 1053–1069. DOI: https://doi.org/10.1016/j.lingua.2012.02.007

Bondevik, Ingrid & Kush, Dave & Lohndal, Terje. 2021. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics* 44(3). 223–254. DOI: https://doi.org/10.1017/S0332586520000207

Bresnan, Joan. 1977. Variables in the theory of transformations. In Culicover, Peter W. & Wasow, Thomas & Akmajian, Adrian (eds.), *Formal syntax*, 157–196. New York, NY: Academic Press.

Brucart, José María. 1993. Sobre la estructura de SCOMP en español. In Viana, Amadeu (ed.), *Sintaxi. Teoria i perspectives*, 59–102. Lleida: Pagès.

Cheng, Ming-Yen & Hall, Peter. 1998. Calibrating the excess mass and dip tests of modality. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60(3). 579–589.

Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, Noam. 1973. Conditions on Transformations. In Anderson, Stephen & Kiparsky, Paul (eds.), *A Festschrift for Morris Halle*, 232–286. Holt, UK: Rinehart and Winston.

Chomsky, Noam. 1977. On *Wh*-Movement. In Culicover, Peter & Wasow, Thomas & Akmajian, Adrian (eds.), *Formal Syntax*, 71–132. New York: Academic Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Cuneo, Nicole & Goldberg, Adele E. 2023. The discourse functions of grammatical constructions explain an enduring syntactic puzzle. *Cognition* 240. 105563. DOI: https://doi.org/10.1016/j.cognition.2023.105563

Erteschik-Shir, Nomi. 1973. *On the nature of island constraints*. Cambridge, MA: MIT dissertation.
Friedmann, Naama & Belletti, Adriana & Rizzi, Luigi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119(1). 67–88. DOI: https://doi.org/10.1016/j.lingua.2008.09.002
Goldberg, Adele E. 2005. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199268511.001.0001
Goldberg, Adele E. 2013. Backgrounded constituents cannot be "extracted." In Sprouse, Jon & Hornstein, Norbert (eds.), *Experimental syntax and island effects*, 221–238. Cambridge, UK: Cambridge University Press.
Hall, Peter & York, Matthew. 2001. On the calibration of Silverman's test for multimodality. *Statistica Sinica* 11(2). 515–536. URL: https://www.jstor.org/stable/24306875
Hartigan, John A. & Hartigan, Pamela M. 1985. The dip test of unimodality. *The Annals of Statistics* 13(1). 70–84. URL: https://www.jstor.org/stable/2241144
Hernanz Carbó, María Lluïsa. 2012. Sobre la periferia izquierda y el movimiento: el complementante "si" en español. In Brucart, José María & Gallego, Ángel J. (eds.), *El movimiento de constituyentes*, 151–171. Madrid: Visor.
Hofmeister, Philip & Sag, Ivan A. 2010. Cognitive constraints and island effects. *Language* 86(2). 366–415. DOI: https://doi.org/10.1353/lan.0.0223
Keshev, Maayan & Meltzer-Asscher, Aya. 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 & Kutas, Marta. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8(4). 573–633. DOI: https://doi.org/10.1080/01690969308407588
Kobzeva, A., Sant, C., Robbins, P. T., Vos, M., Lohndal, T., & Kush, D. (2022). Comparing Island Effects for Different Dependency Types in Norwegian. *Languages*, *7*(3), 197. DOI: http://dx.doi.org/10.3390/languages7030197
Kush, Dave & Lohndal, Terje & Sprouse, Jon. 2018. Investigating variation in island effects: A case study of Norwegian *wh*-extraction. *Natural Language and Linguistic Theory* 36(3). 743–779. DOI: https://doi.org/10.1007/s11049-017-9390-z
Kush, Dave & Lohndal, Terje & Sprouse, Jon. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95(3). 393–420. DOI: https://doi.org/10.1353/lan.2019.0051
Kuznetsova, Alexandra & Brockhoff, Per B. & Christensen, Rune H.B. 2017. lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: https://doi.org/10.18637/jss.v082.i13.
López-Sancio, Sergio. 2015. *Testing syntactic islands in Spanish*. Vitoria-Gasteiz: Universidad of the Basque Country master thesis.
Michel, Dan. 2014. *Individual cognitive measures and working memory accounts of syntactic island phenomena*. San Diego, CA: University of California dissertation.
Morey, Richard D. & Rouder, Jeffrey N. & Jamil, Tahira & Urbanek, Simon & Forner, Karl & Ly, Alexander. 2022. BayesFactor: Computation of Bayes

Factors for common designs (Version 0.9.12-4.5). Retrieved from https://cran.r-project.org/web//packages//BayesFactor/BayesFactor.pdf
Müller, Dietrich Werner & Sawitzki, Günther. 1991. Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association* 86(415). 738–746. DOI: https://doi.org/10.2307/2290406
Ortega-Santos, Iván. 2011. On Relativized Minimality, memory and cue-based parsing. *Iberia: An International Journal of Theoretical Linguistics* 3(1). 35–64.
Ortega-Santos, Iván. 2020. Dialect distance and data assessment in Chilean, Venezuelan and Puerto Rican Spanish. In Rogers, Brandon M.A. & Figueroa Candia, Mauricio A. (eds.), *Lingüística del castellano chileno / Chilean Spanish Linguistics: Estudios sobre variación, innovación, contacto e identidad / Studies on variation, innovation, contact, and identity*, 151–171. Wilmington, DE: Vernon Press.
Ortega-Santos, Iván & Reglero, Lara & Franco, Jon. 2018. *Wh*-islands in L2 Spanish and L2 English: Between poverty of the stimulus and data assessment. *Fontes Linguae Vasconum* 126. 435–471.
Pañeda, Claudia & Kush, Dave. 2022. Spanish embedded question island effects revisited: an experimental study. *Linguistics* 60(2). 463–504. DOI: https://doi.org/10.1515/ling-2020-0110
Pañeda, Claudia & Lago, Sol & Vares, Elena & Veríssimo, João & Felser, Claudia. 2020. Island effects in Spanish comprehension. *Glossa: A Journal of General Linguistics* 5(1). 21. DOI: https://doi.org/10.5334/gjgl.1058
Perlmutter, David M. 1968. *Deep and surface structure constraints in syntax*. New York, NY: Holt, Rinehart and Winston.
Pesetsky, David. 1987. *Wh*-in-situ: Movement and unselective binding. In Reuland, Eric J. & ter Meulen, Alice G. B. (eds.), The representation of (in)definiteness, 98–129. Cambridge, MA: MIT Press.
Pham, Catherine & Covey, Lauren & Gabriele, Alison & Aldosari, Saad & Fiorentino, Robert. 2020. Investigating the relationship between individual differences and island sensitivity. *Glossa: A Journal of General Linguistics* 5(1). 1–17. DOI: https://doi.org/10.5334/gjgl.1199
Prolific. 2023. London, UK. Retrieved from https://www.prolific.co
Qualtrics. 2023. Provo, UT. Retrieved from https://www.qualtrics.com
R Core Team. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
Rizzi, Luigi. 1982. Violations of the *wh*- island constraint and the Subjacency condition. In *Issues in Italian syntax*, 49–76. Dordrecht, Netherlands: Foris.
Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: The MIT Press.
Rizzi, Luigi. 2013. Locality. *Lingua* 130. 169–186. DOI: https://doi.org/10.1016/J.LINGUA.2012.12.002
Rodríguez, Alejandro & Goodall, Grant. 2020. On the universality of *wh*-islands: Experimental evidence from Spanish. In *50th Linguistic Symposium on Romance Languages*, Austin: University of Texas.
Silverman, Bernard W. 1981. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B (Methodological)* 43(1). 97–99. DOI: https://doi.org/10.1111/j.2517-

6161.1981.tb01155.x
Sprouse, Jon. 2007. *A program for experimental syntax: Finding the relationship between acceptability and grammatical knowlege*. College Park, MD: University of Maryland dissertation.
Sprouse, Jon & Caponigro, Ivano & Greco, Ciro & Cecchetto, Carlo. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory Linguistic Theory* 34(1). 307–344. DOI: https://doi.org/10.1007/s11049-015-9286-8
Sprouse, Jon & Fukuda, Shin & Ono, Hajime & Kluender, Robert. 2011. Reverse island effects and the backward search for a licensor in multiple *wh*-questions. *Syntax* 14(2). 179–203. DOI: https://doi.org/10.1111/j.1467-9612.2011.00153.x
Sprouse, Jon & Schütze, Carson T. & Almeida, Diogo. 2013. A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010. *Lingua* 134. 219–248. DOI: https://doi.org/10.1016/j.lingua.2013.07.002
Sprouse, Jon & Villata, Sandra. 2021. Island effects. In Goodall, Grant (ed.), *The Cambridge Handbook of Experimental Syntax*, 227–257. Cambridge, UK: Cambridge University Press.
Sprouse, Jon & Wagers, Matt & Phillips, Colin. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. DOI: https://doi.org/10.1353/lan.2012.0004
Stigliano, Laura & Xiang, Ming. 2021. Experimental evidence on island effects in Spanish relative clauses. *Probus* 33(2). 271–296. DOI: https://doi.org/10.1515/PRBS-2021-0008
Suñer, Margarita. 1991. Indirect questions and the structure of CP: Some consequences. In Campos, Héctor & Martínez-Gil, Fernando (eds.), *Current Studies in Spanish Linguistics*, 283–312. Washington, D.C.: Georgetown University Press.
Szabolcsi, Anna & Zwarts, Frans. 1993. Weak islands and an algebraic semantics for scope taking. *Natural language semantics* 1(3). 235–284. URL: https://link.springer.com/article/10.1007/BF00263545
Torrego, Esther. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15(1). 103–129. URL: https://www.jstor.org/stable/4178369
Villata, Sandra & Rizzi, Luigi & Franck, Julie. 2016. Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua* 179. 76–96. DOI: https://doi.org/10.1016/j.lingua.2016.03.004