Syntax

A Journal of Theoretical, Experimental and Interdisciplinary Research

Syntax 2011 DOI: 10.1111/j.1467-9612.2011.00153.x

Reverse Island Effects and the Backward Search for a Licensor in Multiple *Wh*-Questions

Jon Sprouse, Shin Fukuda, Hajime Ono, and Robert Kluender

Abstract. This paper reports a series of formal acceptability-judgment experiments designed to investigate the syntactic properties of a relatively understudied type of *wh*-dependency: multiple *wh*-questions in English. By using a factorial definition of island effects made available only by formal experiments, we report an unpredicted pattern of acceptability that suggests the existence of *reverse island effects* for *whether* and adjunct islands inside of multiple *wh*-questions, but not for subject and CNPC islands. We argue that this unpredicted effect can best be analyzed by taking into account the parsing processes that are necessary for real-time comprehension of multiple *wh*-questions in English. We propose that multiple *wh*-questions require a *backward search* for an antecedent that is in many ways similar to the *forward search* for a gap site that occurs in single *wh*-questions in English (Frazier & Clifton 1989). We then present additional acceptability-judgment experiments in both English and Japanese to test the predictions of the backward-search analysis.

1. Introduction

Formal experimental techniques for the collection of acceptability judgments (factorial designs and formal statistical tests) have given syntacticians new tools for quantifying the relative contribution of different aspects of the language faculty to the acceptability of a sentence. These new tools have led to a new set of leading questions that have subtly expanded the normal domain of syntactic investigation. For example, the ability to quantify the relative contribution of both grammatical and extragrammatical factors to acceptability has raised the possibility of a more complete theory of the interaction of offline (i.e., atemporal) grammatical operations with the real-time parsing processes that must be deployed to build the structures in question (Frazier 1978; Kluender 1991; Kluender & Kutas 1993b; Phillips 1996, 2003; Fanselow & Frisch 2004; Featherston 2005; Alexopoulou & Keller 2007; Hofmeister 2007; Sprouse 2007; Wagers 2008). *Wh*-dependencies have long been an ideal case study for the interaction of grammatical operations and parsing processes because both types of cognitive operations are relatively well understood in their respective literatures. This paper extends this previous work by applying formal experimental techniques to the investigation of the syntactic properties of a relatively understudied type of *wh*-dependency: multiple *wh*-questions in English. By using a factorial definition of island effects made available only by formal experiments, we report an unpredicted pattern of acceptability that suggests the existence of *reverse island effects* with multiple *wh*-questions. We argue that this unpredicted phenomenon can best be analyzed by taking into account the parsing procedures that are necessary for real-time comprehension of multiple *wh*-questions in English. Specifically, we

This work was supported in part by National Science Foundation grant BCS-0843896 to Jon Sprouse. We would like to thank Norbert Hornstein and two anonymous reviewers for helpful comments on an earlier draft. Any remaining errors are entirely our own.

propose that multiple *wh*-questions require a *backward search* for an antecedent that is in many ways similar to the *forward search* for a gap site that occurs in single *wh*-questions in English (Frazier & Clifton 1989). We then present additional experiments in both English and Japanese to test the predictions of the backward-search analysis.

2. The Syntactic Properties of Multiple *Wh*-Questions

It is well known that *wh*-phrases in *wh*-questions must be assigned scope in order for the question to receive an interpretation (Groenendijk & Stokhof 1997). In the case of single *wh*-questions in English, the scope of the *wh*-phrase is obvious: the scope is isomorphic with the syntactic position of the *wh*-phrase, which must be a specifier of CP:

(1) $[_{CP}$ What did $[_{IP}$ John buy __]]?

For many forms of generative grammar, it is assumed that a grammatical operation called *movement* is responsible for licensing the displacement of the *wh*-phrase that marks its scope. The situation is much more complicated for multiple *wh*-questions in English, as there are (at least) two *wh*-phrases in the sentence (as the name suggests), but only one of those *wh*-phrases appears (potentially displaced by movement) in a specifier of CP.

(2) $[_{CP}$ Who $[_{IP}$ __ bought what]]?

The in-situ *wh*-phrase, like all *wh*-phrases, must be assigned scope in order for the question to receive an interpretation. The empirical fact seems to be that the in-situ *wh*-phrase receives the same scope as the displaced *wh*-phrase (and if there is more than one *wh*-phrase in the specifier of CP, the scope of the in-situ *wh*-phrase is ambiguous). The syntactic question then, is what grammatical operation licenses this scope?

Several different operations have been proposed to give the in-situ *wh*-phrase the same scope as the displaced *wh*-phrase. One early proposal (Huang 1982) assumed a parallelism between the grammatical operation at work for most *wh*-dependencies, namely movement, and the grammatical operation at work for the in-situ *wh*-phrases. This approach postulated a version of movement called *covert movement* that had no overt word order consequences. The benefit of this approach was that it captured the scopal requirements of multiple *wh*-questions with a minimal expansion of the number of grammatical operations necessary to account for the facts. However, the covert movement approach is not without its complications. For example, it is widely assumed that the movement operation is constrained by island constraints (Ross 1967), as can be seen with single *wh*-questions:

(3) a. *Whether* island: *What do you wonder whether John bought __ ?
b. CNPC island: *What did you make the claim that John bought __?
c. Subject island: *What do you think the joke about __ is funny?
d. Adjunct island: *What do you laugh if John buys __?

The problem is that multiple *wh*-questions do not appear to be constrained by island constraints:

(4) a. *Whether* island: Who wonders whether John bought what?
b. CNPC island: Who made the claim that John bought what?
c. Subject island: Who thinks the joke about what is funny?
d. Adjunct island: Who laughs if John buys what?

This fact led Huang (1982) to suggest that covert movement is not constrained by island constraints (or more specifically, by the Subjacency Condition; see Chomsky 1973), thus breaking some of the parallel with overt movement (see also Nishigauchi 1990).

Given the difference between single and multiple *wh*-questions in English with respect to island constraints, some authors have abandoned the overt/covert movement approach entirely. For example, many authors have observed that English multiple *wh*-questions may have more in common with Japanese single *wh*-questions than they do with English single *wh*-questions, as Japanese single *wh*-questions crucially involve an in-situ *wh*-phrase (5a), and do not show island effects (5b).

(5) a. Satoko-ga Shingo-ga nani-o shita-to itta-no?
Satoko-NOM Shingo-NOM what-ACC did-C said-Q
'What did Satoko say that Shingo did __?'
b. Satoko-ga Shingo-ga nani-o shita-kadooka kiita-no?
Satoko-NOM Shingo-NOM what-ACC did-WHETHER asked-Q
'What did Satoko ask whether Shingo did __?'

The empirical facts seem to be that the scope of *wh*-phrases in Japanese questions is determined by the syntactic position of the interrogative particle (either *ka* or *no*). One popular approach suggests that the grammatical operation that establishes the scope relationship between the *wh*-phrase and the interrogative particle is not movement at all, but rather an operation called *unselective binding*, which as the name suggests, binds the interrogative particle and the *wh*-phrase together (Pesetsky 1987, Aoun & Li 1993, Cole & Hermon 1994, Tsai 1994, Hagstrom 1998). This approach argues that the fact that Japanese questions do not show island effects (Lasnik & Saito 1984) is because the grammatical operation movement is constrained by islands, whereas unselective binding is not. Under this approach, English multiple *wh*-questions may use unselective binding to establish the scope of the in-situ *wh*-phrases, thus explaining the lack of island effects in English as well (see also Reinhart 1997 for a similar approach using choice functions).

Given that much of the debate about the grammatical operation underlying multiple *wh*-questions in English is predicated on the presence or absence of island effects, and given that the major benefit afforded by formal experiments is empirical, as a first step it seems reasonable to reconsider the precise empirical claims regarding the absence of island effects in English multiple *wh*-questions. The first judgments were reported by Huang (1982) as exemplified in the following pairwise comparison:

(6) a. *What do you wonder whether John bought __?
    b. Who wonders whether John bought what?

Though the contrast between (6a) and (6b) has been well established for decades, it is also clear that this contrast does not necessarily indicate that there is absolutely no island effect whatsoever with multiple *wh*-questions. It is possible that there is a relatively small island effect that is simply not salient enough to make the sentence as unacceptable as island effects in single *wh*-questions (cf. Fodor 1983, Ross 1987). This ambiguity in the interpretation of (5a) versus (5b) is a prime example of the benefit of the factorial designs licensed by formal experiments: using a common factorial definition of island effects (Kluender & Kutas 1993b; Sprouse, Wagers & Phillips 2010), it is possible to look for a small island effect with multiple *wh*-questions.

The factorial definition of island effects is predicated on two factors, each with two levels; see Table 1. It is perhaps simplest to begin with the basic case of single *wh*-questions. The first factor in this definition is LENGTH, which has two levels: short and long. The second factor is STRUCTURE, again with two levels: nonisland and island. The levels of each factor can be crossed to create four sentences, each with a unique combination of the factor levels (e.g., a 2 × 2 factorial design).

It is easy to see that condition 1 is the baseline condition. This condition contains neither a long-distance dependency nor an island structure, and is therefore likely to receive the highest rating in a judgment survey. Condition 2 adds one component to the baseline: a long-distance *wh*-dependency. This long-distance dependency will likely cause a small decrease in acceptability, as it is well known that long-distance dependencies are more difficult to process than shorter dependencies (Frazier & Clifton 1989; Kluender & Kutas 1993b; King & Kutas 1995; Fiebach, Schlesewsky & Friederici 2002; Phillips, Kazanina & Abada 2005). Condition 3 also adds only one component to the baseline: an island structure. Again, this island structure will likely cause a small decrease in acceptability, at least in the case of *whether* islands, as

Table 1. A factorial island definition for single *wh*-questions

	LENGTH	STRUCTURE	Example
1	Short	Nonisland	Who __ thinks that John bought a car?
2	Long	Nonisland	What do you think that John bought __?
3	Short	Island	Who __ wonders whether John bought a car?
4	Long	Island	What do you wonder whether John bought __?

semantic complexity is well known to affect acceptability. Condition 4 combines both of these components: it contains a long-distance dependency and an island structure. We know a posteriori that this condition will be rated very low because condition 4 is considered ungrammatical.

The factorial definition of island effects allows us to easily distinguish between the presence of an island effect and the absence of an island effect, based solely on the relative ratings of the four conditions in the definition. If an island effect is present, we expect to see a statistically significant super-additive interaction of LENGTH × STRUCTURE, which can be graphically depicted as the nonparallel lines in the left panel of Figure 1. If no island effect is present, then we expect to simply see two main effects of the factors, which can be graphically depicted as the parallel lines in the right panel of Figure 1.

By expanding this definition to multiple *wh*-questions, we can circumvent the logical problem evident in example (5) and determine if there are small island effects for multiple *wh*-questions. This only requires one small change: instead of the factor LENGTH, the multiple *wh*-question definition involves the factor WH, again with two levels. See Table 2.

Just as before, the presence or absence of an island effect will be revealed by the presence or absence of a super-additive interaction, as schematized in Figure 1. Section 3 reports the results of just such an experiment, which will serve as a springboard for the discussion of grammatical operations and parsing processes in section 4.

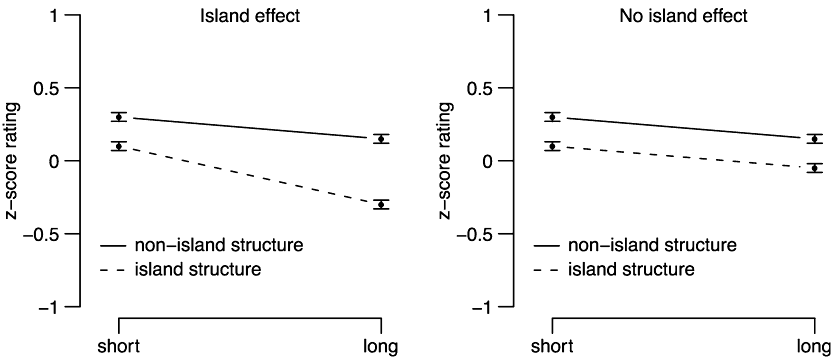

Figure 1. The potential results of the factorial definition of island effects

Table 2. A factorial island definition for multiple *wh*-questions

	WH	STRUCTURE	Example
1	Single	Nonisland	Who __ thinks that John bought a car?
2	Double	Nonisland	Who thinks that John bought what?
3	Single	Island	Who __ wonders whether John bought a car?
4	Double	Island	Who wonders whether John bought what?

3. Experiment 1: Island Effects with Multiple *Wh*-Questions

3.1 Participants

Forty-seven self-reported native speakers of English participated in the experiment through the Amazon Mechanical Turk website (Sprouse 2011). Participants were paid $3 for their participation.

3.2 Materials

Four island types were tested in this experiment: *whether* islands, complex-NP (CNPC) islands, subject islands, and adjunct islands. Each island subdesign was tested in both single *wh*-questions and multiple *wh*-questions. Because two of the conditions overlap between the factorial definition of island effects for single *wh*-questions and multiple *wh*-questions (see Tables 1 and 2), each island sub-design consisted of six conditions. Six example conditions for each island are presented in Table 3.

Twelve lexicalizations of each condition were created and distributed among 6 lists using a Latin square design. This ensured that each participant rated 2 tokens of each condition but never saw related lexicalizations within or across conditions. The 48

Table 3. Examples of each condition for the four islands tested

Island	#	Example
Whether	1	Who thinks that Paul took the necklace?
	2	What does the detective think that Paul took?
	3	Who wonders whether Paul took the necklace?
	4	What does the detective wonder whether Paul took?
	5	Who thinks that Paul took what?
	6	Who wonders whether Paul took what?
CNPC	1	Who heard that Jeff baked a pie?
	2	What did the chef hear that Jeff baked?
	3	Who heard the statement that Jeff baked a pie?
	4	What did the chef hear the statement that Jeff baked?
	5	Who heard that Jeff baked what?
	6	Who heard the statement that Jeff baked what?
Subject	1	Who thinks the gift prompted the congressional hearing?
	2	What does the reporter think prompted the congressional hearing?
	3	Who thinks the gift from the lobbyist prompted the congressional hearing?
	4	Who does the reporter think the gift from prompted the congressional hearing?
	5	Who thinks what prompted the congressional hearing?
	6	Who thinks the gift of what prompted the congressional hearing?
Adjunct	1	Who thinks that the lawyer forgot his briefcase at the office?
	2	What do you think that the lawyer forgot at the office?
	3	Who worries if the lawyer forgets his briefcase at the office?
	4	What do you worry if the lawyer forgets at the office?
	5	Who thinks that the lawyer forgot what at the office?
	6	Who worries if the lawyer forgets what at the office?

target tokens (4 islands × 6 conditions × 2 tokens) in each list were combined with 48 unrelated filler items. The filler items were chosen such that the composition of each survey was 50% acceptable and 50% unacceptable (by hypothesis), as well as 50% interrogative and 50% declarative.

3.3 Presentation

The 96 items in each list were pseudo-randomized such that two related conditions (i.e., conditions from the same island subdesign) never appeared consecutively. Nine additional "anchoring" items (three each of acceptable, unacceptable, and moderate acceptability) were placed as the first nine items of each survey. Participants rated these items just like the others; they were not marked as distinct from the rest of the survey in any way. However, these items were not included in the analysis as they served simply to expose each participant to the full range of acceptability prior to rating the experimental items (a type of unannounced "practice"). The surveys were advertised on the Amazon Mechanical Turk website (see Sprouse 2011 for evidence of the reliability of data collected using AMT), and presented as web-based surveys. Participants completed the surveys using a web browser at their own pace.

3.4 Task

The task was magnitude estimation (Stevens 1957; Bard, Robertson & Sorace 1996). In the magnitude estimation task, participants are presented with a reference sentence, called the *standard*, which is preassigned an acceptability rating, called the *modulus*. Participants are asked to use the standard to estimate the acceptability of the experimental items. For example, if the standard is assigned a modulus of 100, and the participant believes that an experimental item is twice as acceptable as the standard, the participant would rate the experimental item as 200. If a participant believes the experimental item is half as acceptable as the standard, she would rate the experimental item as 50. The standard sentence was in the middle range of acceptability: *Who said that my brother was kept tabs on by the FBI?* The standard was assigned a modulus of 100 and repeated every seven items to ensure that it was always visible on the screen.

3.5 Results

Acceptability judgments from each participant were z-score transformed prior to analysis to eliminate some of the forms of scale bias that potentially arise with scaling tasks. Figure 2 reports the mean z-score rating and standard error of each condition for the single *wh*-question island effects.

To look for statistical interactions indicative of island effects, we ran linear mixed effects models with items and participants included as random factors on each of the island types using LENGTH and STRUCTURE as fixed factors. These linear mixed effects models are comparable to a repeated-measures two-way ANOVA, but with participants and items entering the model simultaneously rather than as separate

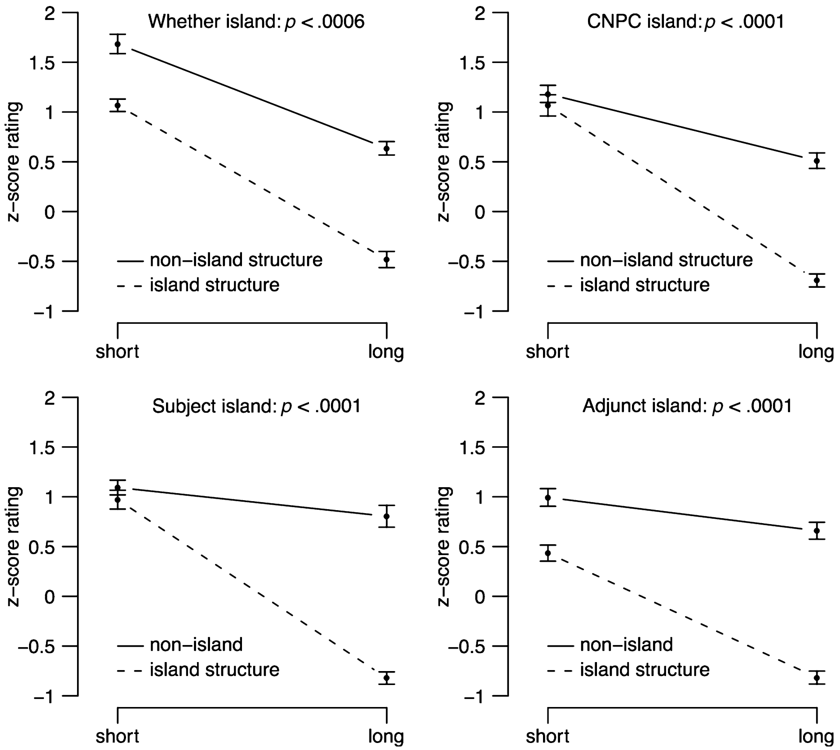

Figure 2. Island effects with English single *wh*-questions. The *p*-values of the interaction term appear at the top.

analyses. Fixed factors were centered prior to analysis, which can help eliminate any potential collinearity between the factors. All *p*-values were estimated using the MCMC method implemented in the languageR package for R (Baayen 2007; Baayen, Davidson & Bates 2008). Table 4 reports the *p*-values for each factor and the interaction for each of the island subdesigns.

As expected based on the existing literature, the linear mixed effects models reveal significant interactions that are indicative of island effects for all four island types in single *wh*-questions.

Table 4. Experiment 1, *p*-values for each term in the two-way linear mixed effects models for each island type in single *wh*-questions (n = 47). Significant effects at $p < .05$ are marked with an asterisk.

	Whether	Complex NP	Subject	Adjunct
LENGTH	.0001*	.0001*	.0001*	.0001*
STRUCTURE	.0001*	.0001*	.0001*	.0001*
LENGTH × STRUCTURE	.0006*	.0001*	.0001*	.0001*

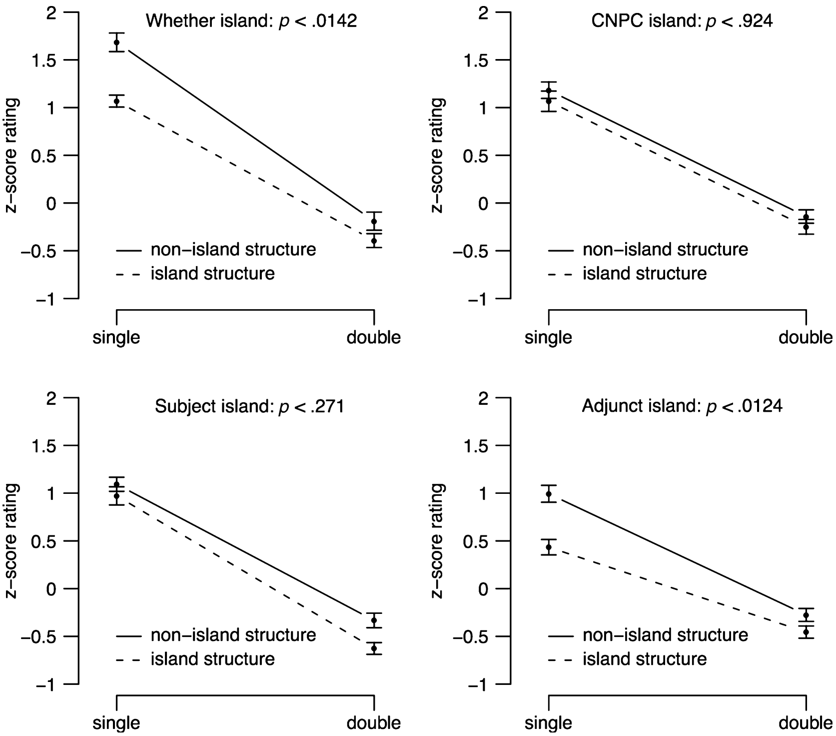

Figure 3. Island effects with English multiple *wh*-questions. The *p*-values of the interaction term appear at the top.

Figure 3 reports the mean z-score rating and standard error of each condition for the multiple *wh*-question island effects. Similar to the analysis of single *wh*-questions, to look for statistical interactions that are indicative of island effects with multiple *wh*-questions, we ran linear mixed effects models with items and participants included as random factors on each of the island types using WH and STRUCTURE as fixed factors. Fixed factors were centered prior to analysis to help eliminate any potential collinearity between the factors. All *p*-values were estimated using the MCMC method implemented in the languageR package for R (Baayen 2007; Baayen, Davidson, & Bates 2008). Table 5 reports the *p*-values for each factor and the interaction for each of the island subdesigns.

Contrary to the expectations established in previous literature, these linear mixed effects models reveal significant interactions for two of the island types in multiple *wh*-questions. However, the graphs in Figure 3 make it clear that the interactions for *whether* and adjunct islands are in fact *sub-additive* rather than *super-additive*: condition 4 in each subdesign is actually more acceptable than would be expected based on the size of the difference in acceptability between conditions 1 and 3 (i.e., the difference between 1 and 3 is larger than the difference between 2 and 4). As this

Table 5. Experiment 1, *p*-values for each term in the two-way linear mixed effects models for each island type in multiple *wh*-questions (*n* = 47). Significant effects at *p* < .05 are marked with an asterisk.

	Whether	Complex NP	Subject	Adjunct
WH	.0001*	.0001*	.0001*	.0001*
STRUCTURE	.0001*	.173	.0062*	.0001*
LENGTH × STRUCTURE	.0142*	.945	.271	.0124*

sub-additive effect in fact runs in the opposite direction to the predicted super-additive definition of island effects, we will refer to this as a *reverse island effect*.

3.6 Discussion

From the point of view of the existing literature, the lack of island effects with CNPC and subject islands is a formal corroboration of the informal results reported by Huang (1982). The lack of island effects reinforces the parallelism between Japanese single *wh*-questions and English multiple *wh*-questions. However, the sub-additive interactions for *whether* and adjunct islands are entirely unpredicted, as they suggest the existence of a *reverse island effect*: an unexpected increase in acceptability when the in-situ *wh*-phrase appears inside a *whether*-island or adjunct-island structure. The question, then, is what could be causing this unexpected increase in the acceptability of multiple *wh*-questions involving in-situ *wh*-phrases that occur within *whether* and adjunct islands.

4. The Parsing Processes Required by Multiple *Wh*-Questions

Recall from section 2 that the lack of island effects with multiple *wh*-questions was interpreted as evidence that the grammatical operation underlying scope assignment in multiple *wh*-questions was not standard movement. To maintain the parallelism with single *wh*-questions, some researchers proposed a new type of movement called *covert movement* that was not constrained by island effects. Other researchers argued that multiple *wh*-questions are more likely parallel to Japanese single *wh*-questions, and therefore suggested novel grammatical operations to assign scope in both Japanese single *wh*-questions and English multiple *wh*-questions, such as unselective binding (Pesetsky 1987, Aoun & Li 1993, Cole & Hermon 1994, Tsai 1994, Hagstrom 1998) and choice functions (Reinhart 1997). The factorial definition of island effects employed in experiment 1 did indeed corroborate the lack of CNPC and subject island effects reported by Huang (1982), and as such reinforces the observed parallelism between English multiple *wh*-questions and Japanese single *wh*-questions. However, neither covert movement, unselective binding, nor choice function operations predict the reverse island effects that occur with *whether* islands and adjunct islands. In this section, we argue that the analysis of reverse island effects requires expanding the domain of inquiry beyond grammatical operations and considering the operations involved in parsing multiple *wh*-questions.

4.1 The Parsing of Single Wh-Questions in English and Japanese

Perhaps the most salient characteristic of English single *wh*-questions is that the *wh*-phrase is displaced (via the grammatical operation movement) from the location where it is (compositionally) semantically interpreted, and instead appears in its scope position, Spec,CP. In the sentence-processing literature, these dependencies are commonly referred to as filler-gap dependencies (Fodor 1978): the displaced *wh*-phrase is the filler, and the base position is the gap. The fact that sentences are produced, and thus comprehended, sequentially (i.e., one word at a time), conspires with the fact that the filler occurs before the gap in English syntax to produce an interesting array of parsing properties. First, because the filler cannot be (compositionally) semantically interpreted in the location that it appears, Spec,CP, its semantic content (and in languages with overt case marking, its morphosyntactic content) must be encoded in memory until the gap location is encountered (e.g., Kluender & Kutas 1993a,b; King & Kutas 1995; Fiebach, Schlesewsky & Friederici 2002; Phillips, Kazanina & Abada 2005). Second, the gap location itself must be identified so that semantic integration can proceed (e.g., Fodor 1978, Crain & Fodor 1985, Stowe 1986, Frazier & Clifton 1989). Third, a set of processes must retrieve the filler from memory and integrate it into the structure at the gap location (e.g., Kaan et al. 2000; Fiebach, Schlesewsky & Friederici 2002; Phillips, Kazanina & Abada 2005).

Although all three of the (sets of) processes deployed during the processing of single *wh*-questions are relevant, the process of identifying the gap location is of particular interest to our investigation. It is well known that this search is *active*: the parser uses word-category information (e.g., that verbs and prepositions are argument-taking categories) to predict possible gap locations before unambiguous confirmation of the gap location is possible (Fodor 1978, Crain & Fodor 1985, Stowe 1986). This active gap-prediction is followed by active retrieval and integration of the filler. Frazier & Flores d'Arcais (1989) and Frazier & Clifton. (1989) famously formulated this as the *active filler strategy*: the parser attempts to complete a filler-gap dependency as early as possible. Classic evidence for the active filling strategy comes from the filled-gap effect (Crain & Fodor 1985, Stowe 1986) and the plausibility effect (Tanenhaus et al 1989), which trigger slow-downs in reading times (and N400 effects; Garnsey, Tanenhaus & Chapman 1989) associated with the incorrect association of the filler with the first verb encountered during incremental parsing. In other words, the displaced *wh*-phrases in English single *wh*-questions initiate an active forward search for the gap location.

Unlike English single *wh*-questions, Japanese *wh*-questions do not (necessarily) involve overt displacement of the *wh*-phrase. This represents an interesting dichotomy with English single *wh*-questions: whereas the position of English displaced *wh*-phrases marks interrogative scope but not semantic interpretation, the location of Japanese in-situ *wh*-phrases marks semantic interpretation but not interrogative scope. The scope of a *wh*-phrase in Japanese is marked by the syntactic position of its associated question particle (*ka* or *no*), which due to the syntactic properties of Japanese necessarily appears attached to a clause-final verb and therefore after the *wh*-phrase. These properties conspire to create interesting differences between the processes triggered during the parsing of English and Japanese single *wh*-questions. For example, there is no reason to believe that the semantic (or morphosyntactic) content of the *wh*-phrase should be encoded in working memory in Japanese, given that it appears in the exact same position in which it is semantically interpreted. There is, however, reason to believe that the *wh*-phrase initiates a forward search, as the *wh*-phrase is lacking one of its requirements—namely, an associated scope-marking question particle in a subsequent C head position rather than a gap location, as there is no gap in a Japanese *wh*-in-situ question. In fact, there is evidence that this forward search for a scope marker is *active* in Japanese (Miyamoto & Takahashi 2000; Aoshima, Phillips & Weinberg 2004; Ueno & Kluender 2009), which suggests an *active-scope-marking* strategy that is similar in many ways to the active filler strategy in English (and other *wh*-movement languages). There is, however, no need for retrieval and integration processes at the question particle in Japanese as there is at the gap location in English, because the *wh*-phrase never needs to be encoded in working memory. In sum, despite differences in the encoding and retrieval requirements of the *wh*-phrase, the in-situ *wh*-phrases in Japanese single *wh*-questions initiate an active forward search for an associated scope-marking particle similar to the active forward search for a gap location in English single *wh*-questions.

4.2 The Parsing of Multiple Wh-*Questions in English*

The next question is what processes are necessary to parse the in-situ *wh*-phrase in a multiple-*wh* dependency in English successfully (we assume that the first *wh*-phrase in such constructions—i.e., in the matrix subject position—is parsed in the same way as the displaced *wh*-phrase in a single *wh*-question). One of the basic empirical facts of English multiple *wh*-questions is that the in-situ *wh*-phrase must appear in the c-command domain of the displaced *wh*-phrase, presumably for scope-marking reasons. The displaced *wh*-phrase, on the other hand, does not have any particular requirements of its own in relationship to an in-situ *wh*-phrase. This asymmetry in dependency is critical to understanding the parsing requirements of multiple *wh*-questions: the item with the requirements appears *after* its licensor in the sequential, linear sentence processing stream.

Given a configuration in which the licensor appears linearly before the licensee during incremental parsing, there are two a priori methods for the parser to establish the dependency:

1. As the parser encounters the initial *wh*-phrase, the parser could record its scope for later use by the in-situ *wh*-phrase (should an in-situ *wh*-phrase happen to occur later).
2. As the parser encounters the later in-situ *wh*-phrase, the parser could initiate a *backward search* for a licensor by reactivating and searching through previously parsed material.

Table 6. Processes triggered during the parsing of *wh*-dependencies

English single *wh*-questions	Japanese single *wh*-questions	English multiple *wh*-questions
Encode *wh*-phrase in memory	(No encoding requirement)	(No encoding requirement)
Search forward for a gap	Search forward for a scope marker	Search backward for a scope marker
Search backward for a filler		
Integrate the filler	(No integration requirement)	(No integration requirement)

Although we have no empirical evidence against option one, there are at least three reasons to believe that option two is more likely on the right track. First, option one would be relatively inefficient given the relative frequency of single *wh*-questions versus the relatively infrequency of multiple *wh*-questions. Given that the process of scope marking is necessary only when an in-situ *wh*-phrase is present, it seems more efficient to delay or postpone scope-marking processes until there is evidence that the process is truly necessary. In other words, the simplest analysis is pursued until there is evidence that a more complex analysis is required. Second, there are other dependencies in language that have the same directionality as this one, such as NPI licensing and reflexive binding dependencies. Current evidence suggests that these dependencies involve a backward search along the lines of option two (Shao & Neville 1998; Xiang, Dillon & Phillips 2009; Dillon et al 2010). Finally, as we will argue in the next section, option two provides an explanation for the results of experiment 1 whereas option one does not.

Assuming that the in-situ *wh*-phrase in multiple *wh*-questions initiates a backward search for a licensor (a displaced *wh*-phrase in Spec,CP) during incremental sentence processing leads to the ontology of parsing procedures and *wh*-dependencies shown in Table 6.

The next question is how this ontology can be used to explain the pattern of results yielded by experiment 1. We turn to this question in the next subsection.

4.3 Explaining the Results with a Backward Search for a Licensor

Experiment 1 revealed what we have termed "reverse island effects" with *whether* and adjunct islands in multiple *wh*-questions. To account for these properties, we propose the following analysis. The backward search for a scope marker must, by definition, reactivate previously parsed material in working memory in order to locate a previously encountered *wh*-phrase in a Spec,CP position. The reactivation of previously parsed material in working memory is likely to be a relatively costly process for the parsing and working memory systems, and therefore likely to be shaped by various efficiency considerations. It has recently been suggested that memory-recall processes may use linguistic structure as a type of addressing system such that items in syntactic positions that are relevant for the linguistic phenomenon driving the retrieval will be preferentially recalled from memory (McElree, Foraker & Dyer 2003; Wagers 2008; Dillon et al 2010). For the purposes of scope marking in multiple *wh*-questions, this approach to working memory would suggest that Spec,CP positions will be preferentially reactivated during the backward search for a licensor (we leave open here the question of whether Spec,CP is the only position reactivated, or simply given special status; see section 5.3). And if the backward search is truly parallel to the forward searches that occur in English and Japanese single *wh*-questions, then it will likely be an *active* search in that it will look to the first syntactic position that could hold a potential licensor. Taken together, these two parsing strategies (i.e., searching actively for relevant syntactic positions) suggest that the Spec,CP position that marks the leading edge of the *whether* and adjunct island structures will be preferentially reactivated by the backward search, as that position is the first available Spec,CP (i.e., starting at the in-situ *wh*-phrase and moving backward through the sentence).

There is one crucial difference between the first Spec,CP position encountered during the backward search in *whether* and adjunct islands and that encountered during the backward search in CNPC and subject islands: the Spec,CP position in *whether* and adjunct islands is filled with semantically complex elements that may share morphosyntactic features with *wh*-words, whereas the Spec,CP position in CNPC and subject islands is filled with either an overt or null semantically vacuous complementizer:

(7) a. *Whether* island: Who wonders [$_{CP}$ whether John bought what]?
b. Adjunct island: Who laughs [$_{CP}$ if John buys what]?

(8) a. CNPC island: Who made the claim [$_{CP}$ that John bought what]?
b. Subject island: Who thinks [$_{CP}$ __ [$_{NP}$ the joke about what] offended the priest]?

Strictly speaking, none of the embedded CP positions in (7)–(8) contains a scope marker for the in-situ *wh*-phrase. In each sentence, the true scope marker is the *wh*-phrase in matrix Spec,CP. However, the elements in the embedded Spec,CP in *whether* and adjunct islands do share certain morphosyntactic features with *wh*-words: *whether* introduces yes/no questions, so it must share something akin to a +Q feature with *wh*-words (though it is crucially not [+wh]); *if* in this data set introduces a conditional, but it can also introduce yes/no questions similar to *whether.*

The preceding analysis rests on two critical assumptions: (i) that *whether,* interrogative *if*, and conditional *if* form a sort of natural class with respect during sentence processing despite their syntactic and semantic differences; and (ii) that *whether*, interrogative *if*, and conditional *if* can substitute for *wh*-phrases with respect to scope marking in multiple *wh*-questions despite the former lacking the [+wh] feature required by the syntactic rules of English. Although, on the surface, both of these assumptions may appear surprising, there are compelling considerations that lead us to believe that each is on the right track. First, there are curious, little-known, and poorly understood—but nonetheless robust—event-related brain potential (ERP) effects suggesting that interrogative *whether*, interrogative *if*, and conditional *if* are all processed similarly: all three elicit a positive voltage fluctuation peaking at 300 ms post stimulus onset (P300) localized over temporoparietal regions of scalp. Originally, this puzzling but very specific brain response appeared to be peculiar to the lexical item *if* (Kluender 1991) regardless of its function, but recent evidence (Sprouse, in prep.) reveals a brain response to *whether* that is remarkably similar in polarity, peak latency, and scalp distribution. Crucially, this response is not seen in other closed-class items with similar semantic effects (e.g., modals and negation, which both lead to meanings involving possible worlds; see Kluender 1991). Second, much recent research on real-time sentence processing has revealed specific scenarios in which the parser does not follow syntactic rules strictly, at least not initially (Townsend & Bever 2001; Ferreira, Ferraro & Bailey 2002; Ferreira & Patson 2007; Phillips, Wagers & Lau 2011). Although the precise implication of these mismatches between syntactic rules and parsing processes is still a matter of active research (and debate), the fact seems to be that the parser sometimes commits syntactic errors as long as the resulting structure is "good enough" (even though a precise definition of "good enough" remains elusive). Thus in the case of *whether* and adjunct islands, it is possible that active backward search may consider interrogative or interrogative-like elements in the embedded Spec,CP good enough for the purposes of scope marking in order to minimize the amount of previous material that must be reactivated to locate a scope marker.

4.4 Summary

In a question involving overt *wh*-movement, the *wh*-filler and its associated gap are mutually dependent on each other for their licensing. In this case, crossing the factors LENGTH (increasing the distance between filler and gap) and STRUCTURE (embedding the gap in an island-inducing environment) results in a super-additive interaction that is greater than the sum of its parts for all island types tested (Figure 1, left panel, and Figure 2). In an English multiple *wh*-question, on the other hand, the in-situ *wh*-phrase crucially depends on the independently licensed, scope-marking *wh*-phrase for its licensing. In this case, crossing the factors LENGTH (increasing the distance between the scope-marking *wh*-phrase and the in-situ *wh*-phrase) and STRUCTURE (embedding the in-situ *wh*-phrase inside an island) does not result in a super-additive interaction, but in either a purely additive interaction (Figure 1, right panel) for CNPC and subject islands, or a sub-additive interaction less than the sum of its parts for *whether* and adjunct islands (Figure 3). In other words, multiple *wh*-questions containing an in-situ *wh*-phrase in either a *whether* (*Who wonders whether John bought what?*) or an adjunct island (*Who laughs if John buys what?*) turn out to be more acceptable than anticipated, all other things being equal.

We have suggested that this relative increase in expected acceptability of *whether* and adjunct islands reflects a foreshortened backward search through previously parsed material for an appropriate licensor in a preceding Spec,CP position: when *whether* (in a *whether* island) or *if* (in the particular adjunct islands tested in this study) is present in the embedded clause, it can by hypothesis serve as a "good enough" licensor for the in-situ *wh*-phrase, thereby truncating the backward-search process. This proposal is consistent with the existence of remarkably similar idiosyncratic brain responses both to *whether* and to interrogative or conditional *if*.

Under this analysis, the backward search across a single clause required for *Who wonders whether John bought what?* or *Who laughs if John buys what?* proves less costly in terms of LENGTH than the backward search across two clauses required for *Who thinks that John buys/bought what?*, which requires the reactivation of more previously parsed material (both the embedded and main clauses instead of only the embedded clause). However, the factor STRUCTURE (i.e., the presence of an island) still incurs an independent processing cost.

5. Evaluating the Analysis

In section 4 we offered an analysis of the reverse island effects revealed in experiment 1 that was predicated on the parsing processes required by multiple *wh*-questions. Although this analysis is firmly grounded in our current understanding of (i) the parsing of the various types of dependencies in human language, (ii) the memory retrieval processes involved in backward searches for licensors, and (iii) the mismatches that sometimes arise between syntax and the parser, it does make several predictions that must eventually be tested. For example, our account of the sub-additive effects in adjunct islands rests crucially on the presence of *if* in the conditional adjuncts we tested. This predicts that multiple *wh*-questions with an in-situ *wh*-phrase embedded in a temporal or causal adjunct would instead yield purely additive effects (Figure 1, right panel) in an acceptability-judgment study, just like the CNPC and subject island conditions in experiment 1 (Figure 3). In this section we present two additional acceptability-judgment experiments that were designed to test other predictions of this analysis.

5.1 Experiment 2: Distance Effects in Various Syntactic Dependencies Requiring Search Processes

Search processes consume parsing resources as they are initiated, which suggests that the length of the search will be a prime factor affecting the overall difficulty of parsing a given construction. Under the linking hypothesis that parsing difficulty influences acceptability judgments, longer search processes should in general lead to lower acceptability ratings than shorter search processes when all other factors are held constant. It is, for example, well established that longer filler-gap dependencies (as in English single *wh*-questions) lead to lower acceptability ratings than shorter filler-gap dependencies (Frazier & Clifton 1989; Kluender & Kutas 1993b; Phillips, Kazanina & Abada 2005; Sprouse 2007; Sprouse, Wagers & Phillips 2010). In section 4 we suggested that a similar length effect lay at the heart of the reverse island effects in experiment 1: the parser treats *whether* and *if* as good enough for purposes of scope marking, thus shortening the backward search in *whether* and adjunct islands containing in-situ *wh*-phrases but not in CNPC and subject islands.

To further test the hypothesis that search length influences acceptability judgments, experiment 2 included not only conditions designed to replicate the length effects found in experiment 1for filler-gap dependencies (9) and multiple *wh*-questions (10) using “D-linked” *wh*-phrases (Pesetsky 1987), but also conditions designed to test for length effects in binding dependencies, as anaphoric binding dependencies also involve a backward search. Given that, to our knowledge, no one has previously looked for length effects in binding dependencies, we included three types: bound-variable dependencies with quantifier antecedents (11), *wh*-antecedents (12), and NP antecedents (13). In all three dependency types the pronoun was a possessive pronoun in order to avoid the interference of binding constraints on the length manipulation. Both sentences in each pair of conditions were matched for length in words and length in clauses. In the following examples, the elements participating in the dependency are underlined.

(9) a. The editorial speculated that the expert knew which proposal the urban community had accepted __ during the meeting. — Short
b. The expert knew which proposal the editorial speculated that the urban community had accepted __ during the meeting. — Long

(10) a. The editorial speculated that the expert knew which urban community had accepted which proposal during the meeting. — Short
b. The expert knew which columnist speculated that the urban community had accepted which proposal during the meeting. — Long

(11) a. I was relieved that the mayor knew if every soldier already received their award. — Short
b. I knew if every soldier was relieved that the mayor already presented their award. — Long

(12) a. I was relieved that the mayor knew which soldiers already received their award. — Short
b. I knew which soldiers were relieved that the mayor already presented their award. — Long

(13) a. I was relieved that the mayor knew if Mandy already received her award. — Short
b. I knew if Mandy was relieved that the mayor already presented her award. — Long

These experimental materials were interleaved with those of experiment 1; therefore, all of the experimental details are identical to those described in section 3.

The results for conditions (9)–(13) are presented in Figure 4 (because there is only one *p*-value for pairwise comparisons, these are reported directly in Figure 4, with no additional table). As Figure 4 makes clear, there were again significant length effects for single and multiple *wh*-questions, as well as for all three types of binding dependencies, as expected given the backward-search process that is hypothesized for (anaphoric) binding dependencies (Dillon et al 2010). These results appear to confirm one of the predictions of the backward-search analysis proposed in section 4.

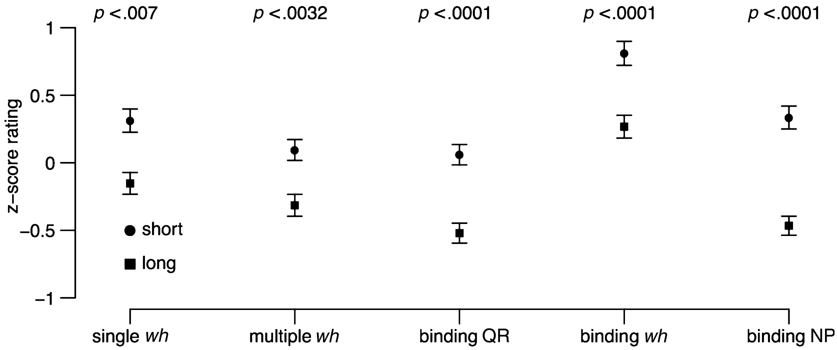

Figure 4. Length effects for constructions requiring search processes in English

5.2 Experiment 3: Island Effects in Japanese Single Wh-*Questions*

Another prediction of the backward-search analysis for multiple *wh*-questions in English is that the extra-sensitive factorial definition of island effects should not reveal reverse island effects with Japanese single *wh*-questions because of the directionality difference between the required searches. The "good enough" licensing behavior of the backward search in English is predicated on the costly reactivation of previously parsed material during a backward search. In contrast, the forward search in Japanese requires no such costly reactivation and, in fact, requires fewer memory resources than the forward search for a gap in English single *wh*-questions because the in-situ *wh*-phrase does not need to be encoded in working memory during the forward search for a scope marker.

To test this prediction, we ran an acceptability-judgment experiment at Kansai Gaidai University in Osaka, Japan, to look for reverse island effects using the factorial definition of island effects. Only *wh*-argument questions were tested, as it is well known that *wh*-adjunct questions do indeed elicit classic island effects in Japanese (Lasnik & Saito 1984). Fifty-four undergraduates participated in the experiment. Eight lexicalizations of each condition were created and distributed among 8 lists using a Latin Square design. Participants rated one token of each condition. The 24 target conditions were mixed with 30 unrelated fillers for a total of 54 items in each survey. The task was magnitude estimation. The standard sentence was *Kenji-wa dare-ga ringo-o kattekita-no-o katteni tabeta-no?* ('Did Kenji eat the apple that who bought?'), and it was assigned a modulus of 100. Example sentences for each of the four conditions are given in Table 7.

As before, the raw ratings were z-score transformed, and the factors were centered prior to analysis. The factors LENGTH and STRUCTURE were entered into linear mixed effects models that included items and participants as random factors. Figure 5 and Table 8 report the results of the experiment. As Table 8 indicates, this experiment corroborated the prediction that there should be no reverse island effects for *whether* islands in Japanese single *wh*-questions (and of course the experiment also corroborated previous results suggesting that there are no classic CNPC or subject

Table 7. Examples of each condition for the four islands tested

Island	#	Example
Whether	1	Dare-ga Shingo-ga ryokoo-o shita-to itta-no? who-NOM Shingo-NOM trip-ACC did-C said-Q ‘Who said that Shingo made the trip?’
	2	Satoko-ga Shingo-ga nani-o shita-to itta-no? Satoko-NOM Shingo-NOM what-ACC did-C said-Q ‘What did Satoko say that Shingo did?’
	3	Dare-ga Shingo-ga ryokoo-o shita-kadooka kiita-no? who-NOM Shingo-NOM trip-ACC did-WHETHER asked-Q ‘Who asked whether Shingo made the trip?’
	4	Satoko-ga Shingo-ga nani-o shita-kadooka kiita-no? Satoko-NOM Shingo-NOM what-ACC did-WHETHER asked-Q ‘What did Satoko ask whether Shingo did?’’
CNPC	1	Dare-ga Satoshi-ga shigoto-o kaeru-to yososhita-no? who-NOM Satoshi-NOM job-ACC change-C predict-Q ‘Who predicted that Satoshi would change his job?’
	2	Kanako-ga Satoshi-ga nani-o kaeru-to yososhita-no? Kanako-NOM Satoshi-NOM what-ACC change-C predict-Q ‘What did Kanako predict that Satoshi would change?’
	3	Dare-ga Satoshi-ga shigoto-o kaeru-toiu yoso-o tateta-no? who-NOM Satoshi-NOM job-ACC change-C prediction-ACC made-Q ‘Who made the prediction that Satoshi would change his job?’
	4	Kanako-ga Satoshi-ga nani-o kaeru-toiu yoso-o tateta-no? Kanako-NOM Satoshi-NOM what-ACC change-C prediction-ACC made-Q ‘What did Kanako make the prediction that Satoshi would change?’
Subject	1	Seijka-ga kojohaisui-ga nani-o yogoshita-to itta-no? politician-NOM wastewater-NOM what-ACC contaminated-C said-Q ‘What did the politician say that the wastewater contaminated?’
	2	Seijka-ga nani-ga machi-o yogoshita-to itta-no? politician-NOM what-NOM town-ACC contaminated-C said-Q ‘Who did the politician say that contaminated the town?’
	3	Satoru-ga seijika-nitsuiteno-kiji-ga dare-niyoru hantaiundo-o Satoru-NOM politician-about-article-NOM who-by protest-ACC okoshita-to itta-no? caused-C said-Q ‘Who did Saturu say that the article about a politician caused the protest by?’
	4	Satoru-ga dare-nitsuiteno-kiji-ga roodoosha-niyoru hantaiundo-o Satoru-NOM who-about-article-NOM working class-by protest-ACC okoshita-to itta-no? caused-C said-Q ‘Who did Saturu say that the article about caused the protest by the working class?’
Adjunct	1	Dare-ga Takeshi-ga kuruma-o katta-to itta-no? who-NOM Takeshi-NOM car-ACC bought-C say-Q? ‘Who said that Takeshi bought a car?’
	2	Keiko-ga Takeshi-ga nani-o katta-to itta-no? Keiko-NOM Takeshi-NOM what-ACC bought-C say-Q? ‘What did Keiko say that Takeshi bought?’

Table 7. Continued

Island	#	Example				
Adjunct	3	Takeshi-ga	kuruma-o	katta-ra	dare-ga	yorokobu-no?
		Takeshi-NOM	car-ACC	bought-COND	who-NOM	be-happy-Q
		'Who would be happy if Takeshi bought a car?'				
	4	Takeshi-ga	nani-o	katta-ra	Keiko-ga	yorokobu-no?
		Takeshi-NOM	what-ACC	bought-COND	Keiko-NOM	be-happy-Q
		'What would Keiko be happy if Takeshi bought?'				

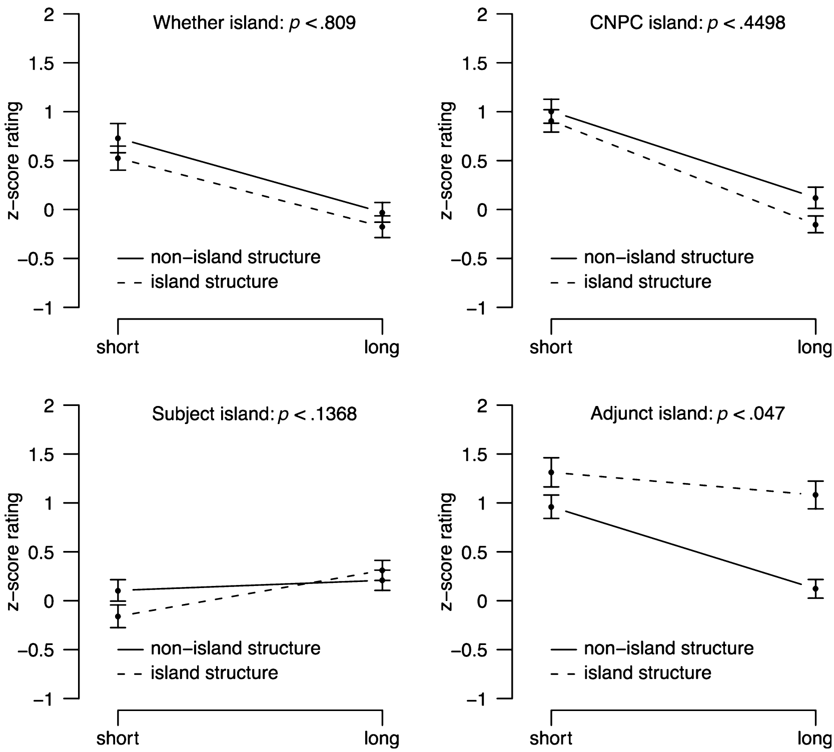

Figure 5. Island effects with Japanese single *wh*-questions. The *p*-values of the interaction term appear at the top.

Table 8. Experiment 3, *p*-values for each term in the two-way linear mixed effects models for each island type (n = 54). Significant effects at $p < .05$ are marked with an asterisk.

	Whether	Complex NP	Subject	Adjunct
WH	.0001*	.0001*	.0204*	.0002*
STRUCTURE	.1642	.1036	.5234	.0001*
LENGTH × STRUCTURE	.809	.4498	.1368	.047*

island effects either). However, there did appear to be a significant interaction of LENGTH × STRUCTURE for adjunct islands. Unfortunately, it is not clear whether this interaction was indicative of a classic island effect or a reverse island effect. The problem appears to be a confound in the adjunct island materials in Japanese. In Japanese, embedded conditional clauses are dispreferred compared to sentence-initial conditional clauses, so the island-structure conditions in Japanese were constructed with sentence-initial conditionals, whereas the nonisland structure conditions (which contain a simple embedded declarative) were left in their canonical order. It seems that the preposing of conditionals in fact causes conditional sentences to be rated better than embedded declaratives: notice that both of the island-structure (conditional) conditions in the adjunct island graph in Figure 5 are rated higher than the nonisland-structure (declarative) conditions. This main effect is reflected in the linear mixed-effect models reported in Table 8 as a main effect of STRUCTURE in adjunct islands. This unexpected reversal in the relative acceptability of conditionals and embedded declaratives makes it impossible to look for true classic island or reverse island effects. However, given that there is indeed no reverese island effect in *whether* islands, it seems likely that there is also no reverse island effect for adjunct islands, either. The yes/no element in *whether* islands (*kadooka*) shares more morphosyntactic features with the question particle *ka* than the conditional marker (*ra*) in adjunct islands. Therefore, *whether* islands seem more likely to reveal potential reverse island effects than adjunct islands.

5.3 Open Questions for Future Research

One primary prediction of the backward-search analysis is that online sentence processing experiments will yield evidence of the backward search during real-time comprehension. Though such experiments are beyond the scope of this article, the speed-accuracy tradeoff (SAT) methodology used by Dillon et al. (2010) to look for evidence of the backward search for an antecedent for reflexives in Mandarin Chinese would likely be a good first step. The backward-search analysis also raises questions about which syntactic positions are reactivated: Is it solely Spec,CP that is reactivated (because Spec,CP is where scope markers appear), or is the rest of the structure reactivated as well? This question could be investigated by looking for "good enough" effects caused by elements in positions other than Spec,CP. Finally, recent ERP evidence has suggested meaningful differences in the pattern of scalp potentials between English single *wh*-questions and Japanese single *wh*-questions (Ueno and Kluender 2009). Of particular relevance to experiment 3, there is no electrophysiological evidence for any kind of backward-search process at the question particle that assigns scope to the in-situ *wh*-phrase. The parallels and differences between the backward search in English multiple *wh*-questions and the forward search in Japanese single *wh*-questions raise questions as to what extent the scalp potentials elicited by the backward search in English multiple *wh*-questions will pattern like Japanese single *wh*-questions, and to what extent they may possibly pattern like the forward search for a gap position in English single *wh*-questions.

6. Conclusion

This paper presented a series of formal acceptability-judgment experiments designed to investigate the syntactic properties of English multiple *wh*-questions, with a focus on previous claims that there are no island effects when the in-situ *wh*-phrase appears within an island structure (Huang 1982). The factorial definition of island effects (Kluender & Kutas 1993b; Sprouse, Wagers & Phillips 2010) corroborated these claims for CNPC and subject islands but also revealed a previously unobserved increase in acceptability within *whether* and adjunct islands that we termed a *reverse island effect*. To account for the reverse island effect, we proposed an analysis based on the parsing processes that are required by multiple *wh*-questions. Specifically, we argued that the in-situ *wh*-phrase in multiple *wh*-questions initiates an *active backward search* through previously parsed material for a viable scope marker (similar to the backward searches proposed for NPIs and reflexive licensing in Xiang, Dillon & Phillips 2009 and Dillon et al. 2010) and that the search for a scope marker considers the elements in Spec,CP of *whether* and adjunct islands (*whether* and *if*, respectively) "good enough" for the purposes of marking the scope of *wh*-in-situ elements (Townsend & Bever 2001; Ferreira, Ferraro & Bailey 2002; Ferreira & Patson 2007; Phillips, Wagers & Lau 2011). We then reported two additional experiments designed to test the predictions of this analysis both in English and Japanese. These experiments corroborated the basic predictions of the analysis, as well as revealing two previously unreported effects: a significant effect of length for binding dependencies and a significant preference for sentence-initial conditionals in Japanese. The backward-search analysis presented here builds on independently motivated components of grammatical and parsing theories and capitalizes on the parallels between multiple *wh*-questions and single *wh*-questions in English and Japanese. In the process of exploring the properties of multiple *wh*-questions, this study has demonstrated the utility of combining experimental syntax techniques with analyses that incorporate both grammatical and parsing theories.

References

Alexopoulou, T. & F. Keller. 2007. Locality, cyclicity and resumption: At the interface between the grammar and the human sentence processor. *Language* 83:110–160.

Aoshima, S., C. Phillips & A. Weinberg. 2004. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51:23–54.

Aoun, J. & A. Li. 1993. *Wh*-elements in situ: Syntax or LF? *Linguistic Inquiry* 24:199–238.

Baayen, R. H. 2007. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., D. J. Davidson & D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390–412.

Bard, E. G., D. Robertson & A. Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72:32–68.

Chomsky, N. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. S. Anderson & P. Kiparsky, 237–286. New York: Holt, Rinehart and Winston.

Cole, P. & H. Hermon. 1994. Is there LF *wh*-movement? *Linguistic Inquiry* 25:239–262.

Crain, S. & J. Fodor. 1985. How can grammars help parsers? In *Natural language parsing: Psycholinguistic, computational, and theoretical approaches*, ed. D. Dowty, L. Karttunen & A. Zwicky, 94–128. Cambridge: Cambridge University Press.

Dillon, B., W. Y. Chow, M. Wagers, T. Guo, F. Liu & C. Phillips. 2010. The structure-sensitivity of search: Evidence from Mandarin Chinese. Ms., University of Maryland, College Park; University of California, Santa Cruz; Beijing Normal University.

Fanselow, G. & S. Frisch. 2004. Effects of processing difficulty on judgments of acceptability. In *Gradience in grammar*, ed. G. Fanselow, C. Féry, M. Schlesewsky & Ralf Vogel. Oxford: Oxford University Press.

Featherston, S. 2005. Universals and grammaticality: *Wh*-constraints in German and English. *Linguistics* 43:667–711.

Ferreira, F., V. Ferraro & K. G. D. Bailey. 2002. Good-enough representations comprehension. *Current Directions in Psychological Science* 11:11–15.

Ferreira, F. & N. Patson. 2007. The "good enough" approach to language comprehension. *Language and Linguistics Compass* 1:71–83.

Fiebach, C. J., M. Schlesewsky & A. D. Friederici. 2002. Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language* 47:250–272.

Fodor, J. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9:427–473.

Fodor, J. 1983. Phrase structure parsing and the island constraints. *Linguistics and Philosophy* 6:163–222.

Frazier, L. 1978. On comprehending sentences: Syntactic parsing strategies. Ph.D. dissertation, University of Connecticut, Storrs.

Frazier, L. & C. Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4:93–126.

Frazier, L. & G. Flores d'Arcais. 1989. Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28:331–344.

Garnsey, S., M. Tanenhaus & R. Chapman. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18:51–60.

Groenendijk, J. & M. Stokhof. 1997. Questions. In *Handbook of logic and language*, ed. J. van Benthem & A. ter Meulen, 1055–1124. Cambridge, MA:MIT Press.

Hagstrom, P. 1998. Decomposing questions. Ph.D. dissertation, MIT, Cambridge, MA.

Hofmeister, P. 2007. Representational complexity and memory retrieval in language comprehension. Ph.D. dissertation, Stanford University, Stanford, CA.

Huang, C.-T. J. 1982. Move WH in a language without WH-movement. *The Linguistic Review* 1:369–416.

Kaan, E., A. Harris, E. Gibson & P. Holcomb. 2000. The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes* 15:159–201.

King, J. & M. Kutas. 1995. Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience* 7:376–395.

Kluender, R. 1991. Cognitive constraints on variables in syntax. Ph.D. dissertation, University of California, San Diego.

Kluender, R. & M. Kutas. 1993a. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5:196–214.

Kluender, R. & M. Kutas. 1993b. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8:573–633.

Lasnik, H. & M. Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15:235–289.

McElree, B., S. Foraker & L. Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48:67–91.

Miyamoto, E. T. & S. Takahashi. 2000. The processing of *wh*-phrases and interrogative complementizers in Japanese. In *Japanese/Korean Linguistics 10*, ed. N. Akatsuka & S. Strauss, 62–75. Stanford, CA:CSLI Publications.

Nishigauchi, T. 1990. *Quantification in the theory of grammar*. Dordrecht: Kluwer.
Pesetsky, D. 1987. *Wh*-in-situ: Movement and unselective binding. In *The representation of (in)definiteness*, ed. E. Reuland & A. ter Meulen, 98–129. Cambridge MA: MIT Press.
Phillips, C. 1996. Order and structure. Ph.D. dissertation, MIT, Cambridge, MA.
Phillips, C. 2003. Linear order and constituency. *Linguistic Inquiry* 34:37–90.
Phillips, C., N. Kazanina & S. Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22:407–428.
Phillips, C., M. Wagers & E. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In *Experiments at the interfaces* (Syntax & Semantics 37), ed. J. Runner. Bingley, UK:Emerald Publications.
Reinhart, T. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20:335–397.
Ross, J. R. 1967. Constraints on variables in syntax. Ph.D. dissertation, MIT, Cambridge, MA.
Ross, J.R. 1987. Islands and syntactic prototypes. In *Papers from the general session at the 23rd annual regional meeting of the Chicago Linguistic Society*, ed. B. Need, E. Schiller & A. Bosch, 309–320. Chicago: Chicago Linguistic Society.
Shao, J. & H. J. Neville. 1998. Analyzing semantic processing using event-related brain potentials. *Newsletter of the Center for Research in Language (University of California, San Diego)* 11:3–20.
Sprouse, J. 2007. A program for experimental syntax. Ph.D. dissertation, University of Maryland, College Park.
Sprouse, J. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43. DOI 10.3758/s13428-010-0039-7.
Sprouse, J., M. Wagers & C. Phillips. 2010. A test of the relation between working memory capacity and syntactic island effects. Ms., University of California, Irvine; University of California, Santa Cruz; University of Maryland, College Park.
Stevens, S. S. 1957. On the psychophysical law. *Psychological Review* 64:153–181.
Stowe, L. 1986. Parsing *wh*-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1:227–245.
Tanenhaus, M. K., J. Boland, S. M. Garnsey & G. Carlson. 1989. Lexical structure in parsing long-distance dependencies. *Journal of Psycholinguistics* 18:37–50.
Townsend, D. & T. Bever. 2001. *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
Tsai, W.-T. D. 1994. On nominal islands and LF extraction in Chinese. *Natural Language & Linguistic Theory* 12:121–175.
Ueno, M. & R. Kluender. 2009. On the processing of Japanese *wh*-questions: An ERP study. *Brain Research* 1290:63–90.
Wagers, M. 2008. The structure of memory meets memory for structure in linguistic cognition. Ph.D. dissertation, University of Maryland, College Park.
Xiang, M., B. Dillon & C. Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* 108:40–55.

Jon Sprouse
University of California, Irvine
Department of Cognitive Sciences
3151 SSPA
Irvine, CA 92697
USA

jsprouse@uci.edu

Shin Fukuda
University of Hawai'i, Mānoa
Department of East Asian Languages and Literatures
Moore Hall 382
1890 East-West Road
Honolulu, HI 96822
USA

fukudash@hawaii.edu

Hajime Ono
Kinki University
Department of Language Education
Osaka
Japan

onohajime@kindai.ac.jp

Robert Kluender
University of California, San Diego
Department of Linguistics
9500 Gilman Drive
La Jolla, CA 92093-0108
USA

kluender@ling.ucsd.edu