Cognitive Constraints on Syntactic Islands

Philip Hofmeister and Ivan A. Sag

Abstract

Competence-based theories of island effects play a central role in generative grammar, yet the graded nature of many syntactic islands (Deane 1991; Kluender 1992) has never been properly accounted for. Indeed, syntactic accounts of island effects have persisted in spite of a wealth of data suggesting that island effects are not categorical in nature and that non-structural manipulations (i.e. ones leaving the island intact) can radically alter judgments of island violations. We argue here that processing (performance) factors have the potential to account for this otherwise unexplained variation in acceptability judgments. Processing factors are known to play a significant role in the perception of (un)acceptability (e.g. Chomsky & Miller, 1963); however, processing-based accounts of variation in acceptability judgments are not well-represented in the literature. Notably, syntactic island constructions are characterized by linguistic properties that are predicted to incur heavy processing costs and which have been noted by previous authors as markedly influencing acceptability judgments.

Self-paced reading experiments and controlled acceptability studies are used here to explore the relationship between processing costs and judgments of acceptability. In each of the three self-paced reading studies, the data indicates that the processing cost of different types of island violations can be significantly reduced to a degree comparable to that of non-island filler-gap constructions by manipulating a single non-structural factor. Moreover, this reduction in processing cost is accompanied by significant improvements in acceptability. This evidence favors the hypothesis that island-violating constructions involve numerous processing pressures that aggregate to drive processing difficulty above a threshold so that a perception of unacceptability ensues. We also consider the extent to which competence-based theories can account for these and related results and the implications of these findings for syntactic theories of filler-gap dependencies.

1 Introduction

It is generally agreed that grammar imposes no upper bound on the distance spanned by filler-gap dependencies in English $wh$-constructions like (1):

(1) What did Sawyer say he thought everyone knew . . . $s$ the president had eaten ___?

In addition, as more than a half century of intense debate has made clear, related examples like (2)–(4) sound unacceptable to most speakers of English:
To explain the deviance of these examples and that of other filler-gap constructions, syntacticians have long supposed the existence of so-called ‘island’ constraints: grammatically imposed restrictions effectively preventing any elements from escaping from certain configurationally defined environments. Since Chomsky (1962), island constraints on filler-gap dependencies have played a major role in the development of modern theories of syntax. Not only have the exegeses surrounding these data occupied a major portion of the central texts of transformational grammar (Chomsky, 1973, 1977, 1981, 1986), but countless other syntactic analyses have depended upon islands as a diagnostic for movement or variable binding.

In spite of their importance in current syntactic theory, island constraints have not enjoyed a stable analysis in the literature; in fact the very data that underlie these constraints have been slippery at best. To varying degrees, syntactic theories of islandhood have recognized a heterogeneity in the universality and rigidity of different island constraints and have attempted to explain various counterexamples on the basis of their being peripheral, allowed via some exception principle, or allowed because they do not constitute the relevant sort of linguistic dependency. Most important for the current discussion is the fact that the acceptability of sentences containing island violations appears to vary systematically with the manipulation of nonstructural factors, i.e. manipulations that leave the phrase structure configuration intact.

The present analysis follows a competing line of thought – that the variance of acceptability judgments associated with these constructions, both language-externally and cross-linguistically, can be better explained by appealing to cognitive constraints on language processing (Deane, 1991; Kluender & Kutas, 1993b; Kluender, 1991, 1992, 1998, 2005; Alexopoulou & Keller, 2007). More generally, this view follows a tradition that analyzes acceptability as the end result of a series of cognitive processes. Accordingly, acceptability reflects the contribution of grammatical principles (competence-based factors), as well as constraints imposed by resource limitations (performance-based factors). Such an approach is articulated by Chomsky and Miller (1963), who argue that the unacceptability of complex center-embedding constructions is the result of memory limitations:

\[(5) \quad \text{The boy the girl the host knew brought left.}\]

This perspective – that processing difficulty sometimes accounts for the perception of ungrammaticality – also underlies the standard treatment of so-called garden path sentences, e.g. The horse raced past the barn fell. In this case, the preference for the matrix verb reading of raced is so strong that the parser encounters significant difficulty in finding the appropriate reduced relative reading and a perception of ungrammaticality ensues (Bever, 1970; Osterhout, Holcomb, Swinney, 1994).
The unacceptability of sentences like (2)–(4) above, in contrast, has not been widely regarded as stemming from limitations on memory or other cognitive resources. The syntactic literature instead has maintained the position that examples like these reflect universal principles of grammar, e.g. the Complex NP Constraint, Wh-Island Constraint, or the Subjacency Condition.

In this paper, we argue that at least some island phenomena, in particular examples like (2)–(4), owe their character to the accumulation of performance-related difficulties that rises above some threshold to create a perception of ungrammaticality. This position is supported by a number of facts that are difficult to reconcile with any theory of island effects based on fixed (i.e. immutable) grammatical constraints. First, many island constructions involve factors known independently to engender processing difficulty. Variation in acceptability both within languages and across languages appears to correlate with the presence of such processing hurdles. Additionally, we present empirical evidence here, from controlled acceptability studies and behavioral measures of processing, which indicates that the same factors that create processing difficulty also lower judgments of acceptability. Indeed, controlling the factors that cause these processing difficulties can dramatically increase the acceptability of island-violating sentences.

It is also a striking fact that counterexamples have been discovered to nearly every proposed island constraint and this fact poses a further serious issue for theories that subscribe to explanation in terms of universal, inviolable constraints of grammar. In contrast, a theory that allows for sentences with the same structural configuration to evoke different processing loads (on the basis of non-structural manipulations) both explains and predicts the existence of such examples, provided they reflect linguistic choices that significantly reduce processing difficulty. To be clear, the existence of counterexamples to proposed grammatical constraints does not necessitate a processing-based reinterpretation of the data; however, in tandem with the kinds of evidence outlined above, these counterexamples are confirmation of the predictions of a processing-driven account.

On the basis of these arguments, we suggest a grammar of filler-gap dependencies (FGDs) that is far simpler than that generally countenanced in the syntactic literature, more ‘minimal’ even than Chomsky’s (1995) ‘Minimalist’ Program. In our view, a competence grammar licenses a superset of the possible filler-gap sentences, leaving it to processing factors to explain why only a proper subset of these grammatically licensed sentences (including those like (2)–(4) above) are fully acceptable.

Our claim that competence grammar has no need for a Subjacency Condition or a Wh-Island Constraint should, of course, not be construed as a claim that there are no constraints (or even universal island constraints) within competence grammar. However, an analysis of these island effects in cognitive terms opens the way to a more homogenous and transparent set of grammatical constraints on FGDs. That is, excluding some of the restrictions on dependency formation from the purview of grammar may, in fact, lead to a clearer understanding of whatever grammatical constraints on dependencies remain, once the burden of grammar is lightened by a deeper understanding of interacting performance factors.

To lay the groundwork for such an analysis, we begin the next section by briefly reviewing the central theories of island phenomena in the syntactic literature. The history of this theorizing itself reveals an implicit acknowledgment of the variance associated with these construc-
tions. Not only are the judgments surrounding the key data often questioned by the authors themselves, but the resulting theories are often forced to reclassify linguistic dependencies in seemingly unmotivated ways in order to preserve the force of the basic generalizations. In the following section, many of the inherent processing difficulties imposed by island constructions are identified and discussed in the context of the discoveries made by psycholinguistic research. This discussion reveals that many island effects are accompanied by a variety of burdens on cognitive resources that are typically not controlled for in syntactic research. While no single processing burden is likely to pose a serious problem for the human sentence processor, the combination of multiple, simultaneous demands and costs can become overwhelming.

After reviewing these factors, we present a series of experiments that evaluate the impact of one variable on the processing and acceptability of island-violating sentences: the syntactic and semantic complexity of the filler-phrase. Each of these experiments shows that increasing the complexity of the filler-phrase significantly facilitates the processing of FGDs, attenuating the island effect. Crucially, these experiments demonstrate that this manipulation effectively eliminates the processing difference between dependencies that cross island boundaries and those that do not. In the remaining sections, we evaluate how both grammar-based theories and processing-based theories fare with respect to the data we discuss. We conclude with a general discussion of the implications of such an approach for theories of syntax.

2 Theories of Islandhood

2.1 A Brief History of Islands

The earliest attempt to lay out a general principle to restrict the set of possible FGDs comes from Chomsky (1962) where the A-over-A Condition is introduced:

(6) The A-over-A Condition (AOAC)

An element of category A cannot be extracted out of a phrase of category A.

This principle consequently rules out sentences where an NP has been extracted from an NP, as in (7a-b):

(7) a. What did he know [someone who has ___]?  
   b. What did you see the man read [the book that was on ___]?

However, as Ross (1967) details, the AOAC predicts ungrammaticality for tokens considered acceptable [(8)–(9)] and grammaticality in cases judged unacceptable [(10)]:

(8) Who would you approve of [NP my seeing ___]?  
(9) Which astronaut did you read [NP a book about ___]?
(10) Which dignitaries do you think [(Sandy photographed the castle) and [Chris visited ___]]?
Counterexamples like these led Ross to introduce a number of distinct island constraints that are still part of the descriptive vocabulary of the modern syntactic literature, such as the Complex Noun Phrase Constraint:

(11) **The Complex NP Constraint (CNPC)**

No element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation.

These individual constraints cover more empirical ground than Chomsky’s original proposal, allowing for examples like (8) and (9), but ruling out examples like (10) on the basis of a separate constraint (the Coordinate Structure Constraint).\(^1\)

Like the AOAC, however, the individual constraints proposed by Ross were not without exceptions, many of which were observed and discussed by Ross himself. For instance, Ross observed that tense significantly alters the acceptability of some island-violating dependencies, e.g. CNPC violations, as well as modality introduced by periphrastic phrases like *make the claim that* or *have hopes that*. Thus, (12a) sounds better than examples like (12b):

(12) a. How much money are you making the claim that the company squandered a large amount of money? ≥

b. How much money are you stating the fact that the company squandered?

According to Ross, this perceived difference did not refute the validity of the proposed principle. Examples like (12a) were instead removed from the domain of relevance because they were assigned a significantly different syntactic structure, one where the *that*-clause functions as a complement of the verb, not the noun.

Chomsky (1973) subsequently proposed a more general principle to account for a range of island phenomena. In this ‘Subjacency’ theory, nestings of certain phrasal categories blocked movement:

(13) **The Subjacency Condition**

No rule may move a phrase from position Y to position X (or conversely) in:

\[
\ldots X \ldots [\alpha \ldots [\beta \ldots Y \ldots] \ldots] \ldots X \ldots
\]

where \(\alpha\) and \(\beta\) are cyclic nodes.

In English, IP and NP (or DP) are cyclic nodes, meaning that complex noun phrases like \([NP the rumor that [IP they started a new company]]\) and subject noun phrases, as in \([IP[NP the attempt to find the fountain of youth] ended in failure]]\), constitute islands for movement. Hence, the grammar excludes examples like the following:

(14) a. It was a new company that Simon spread \([NP the rumor that [IP they started ___]]\).

\(^1\)Although Ross is often criticized for proposing ad hoc constraints of insufficient generality, it should be noted that Ross (1967) includes an attempt to unite many of his proposed constraints under a principle of ‘A directly over A’.
b. What did $[IP_{NP: the attempt to find ___ }]$ end in failure? 

Subjacency accordingly limits how many cyclic nodes a FGD can legally cross. Because of its formulation in terms of cyclic nodes, it also had the potential to explain constraints on superficially disparate structures, giving it a desirable generality.

However, as demonstrated by Deane (1991), it is possible to construct reasonably acceptable examples where three or more cyclic nodes are crossed in English:

(15) a. Nixon was one president that $[IP_{NP: they had [NP: no trouble finding [NP: votes for [NP: the impeachment of ___ ]]}]}$.

b. The chief purpose is $[IP_{NP: to avoid any sentences that [IP_{NP: our informants report [NP: significant variations in [NP: their judgments about ___ ]]}]}]$.

Moreover, Ross explicitly discussed similarly complex data like the following:

(16) Which reports does the government prescribe the height of the lettering on ___?

Despite the fact that such examples of ‘deep extraction’ pose direct counterevidence to the Subjacency Condition, the field has generally ignored them, presumably following Chomsky’s (1973) unsubstantiated assertion that such patterns are ‘not generally possible’ (see Deane 1991, p. 11).

Reasoning much as Ross did in his discussion of make the claim, Chomsky acknowledges that certain examples ruled out by the Subjacency Condition are fully grammatical. He points out, for example, that the choice of clause-initial wh-element seems to affect the grammaticality of the dependency:

(17) a. What crimes does the FBI know how to solve ___?

b. *What crimes does the FBI know whether to solve ___?

Admitting the graded acceptability of extractions out of embedded questions, Chomsky concludes that “the know how to examples such as (59) [(17a) here] are unique in permitting further wh-Movement [of the sort illustrated here]” (Chomsky 1973, p. 245). That is, examples of this type are excluded or regarded as ‘unique’ because the Subjacency theory fails to account for them, rather than any theory-independent considerations. We are not arguing against the possibility of there being something grammatically special or extraordinary about certain phrases or collocations. Our purpose here is rather to point out a consistent pattern in the literature – the pattern of asserting without argument that counterexamples are ‘special’ or ‘exceptional’, thereby rendering them irrelevant for purely grammatical theories of syntactic islandhood.

In another case, where variation across speakers is evident, Chomsky questions the reliability or standardness of some judgments:

“Some speakers seem to accept such forms as What did he wonder whether John saw? What crimes did he wonder how they solved? For me, these are unacceptable. It would be possible to add special rules to allow for these examples by a complication of the particular grammar, given the suggested interpretation of the conditions.”

[Chomsky, 1973, p. 244]
Problematic examples and speakers are again treated as ‘special’. The grammars of speakers who allow such tokens are non-optimal, by implication, since a more complex grammar is necessary to explain their perceptions. The practical outcome of such pronouncements is that these examples do not have to be seriously considered or related to their minimally different counterparts.

Similarly, there is also a very open acknowledgment of ‘the gradation of acceptability’ for extraction of picture NPs, where manipulating definiteness appears to affect acceptability:

“Some speakers (myself included) find a three-way gradation of acceptability, with (30b) [Who did you see pictures of?] better than Who did you see the pictures of?, which is in turn preferable to (31b) [Who did you see John’s picture of?]”

[Chomsky, 1973, p. 239]

At the time, however, the Subjacency theory had no means of expressing this gradation, nor was it predicted by any aspect of the theory. But in a hint of things to come, Chomsky remarks that some structures might involve a ‘double violation’ and others only a ‘single violation’.

The Barriers account (Chomsky, 1986) realizes this notion of gradience in a more explicit fashion. According to this theory, certain XPs act as barriers to movement or extraction, specifically XPs that are not theta-governed (or L-marked, as in the definition below) by a lexical category, i.e. phrases not selected by a governing lexical head.

\[
\gamma \text{ is a BC [blocking category] for } \beta \text{ iff } \gamma \text{ is not L-marked and } \gamma \text{ dominates } \beta
\]

\[
\gamma \text{ is a barrier for } \beta \text{ iff (a) or (b):}
\]

a. \(\gamma\) immediately dominates \(\delta, \delta\) a BC for \(\beta\);
b. \(\gamma\) is a BC for \(\beta, \gamma \neq \text{IP}\).

Unlike the earlier Subjacency theory, the Barriers treatment explicitly acknowledged different degrees of acceptability, determined by the number of barriers crossed by movement:

\[
\beta \text{ is } n\text{-subjacent to } \alpha \text{ iff there are fewer than } n + 1 \text{ barriers for } \beta \text{ that exclude } \alpha.
\]

Dependencies which cross zero boundaries are thus 0-subjacent and should sound perfectly acceptable, ‘1-subjacency’ should translate to marginal acceptability, but anything higher “should yield a considerable decrement in acceptability” (ibid., p. 30).²

As an illustration, examples like (20a-b) involve movement across two barriers, according to Chomsky (1986):

\[
(20) \quad \text{a. What did Simon spread } [NP \text{ the rumor } [CP \text{ that } [IP \text{ they started } \ldots ]]]? \\
\text{b. What did Harold wonder } [CP \text{ whether } [IP \text{ they had ruined } \ldots ]]? \\
\text{c. What did the captain give } [NP \text{ the command } [CP [IP \text{ to start } \ldots ]]]? \\
\text{d. Who did Adele wonder } [CP \text{ whether } [IP \text{ to invite } \ldots ]]?
\]

²Based on the implications of the text, subjacency violations beyond 1-subjacency should be indiscriminable in terms of acceptability, because “to specify n-subjacency for higher values of n requires counters.” In other words, a dependency that cross two barriers should incur the same degradation in acceptability as a dependency that crosses three or four barriers.
One barrier is posed by the CP, which acquires this status since it immediately dominates a blocking category – the most embedded IP. The second barrier is actually the IP itself, despite the exception statement in (18) suggesting that IPs cannot be inherent barriers. This is because, in order to account for the apparent difference between extraction out of a tensed island [as in (20a) and (20b)], Chomsky assumes that the most deeply embedded tensed IP constitutes an inherent barrier to movement. (20c) and (20d) each present only a single barrier to movement: the CP that again earns its barrier status via inheritance from the IP. In these cases, the IP is only a blocking category and not a barrier, since the IP is not tensed. The latter two examples should thus sound better than the first two, if the predictions of the Barriers account are accurate. This theory accordingly recognizes some amount of gradience in the island data, albeit limited to three distinct levels of grammaticality.

Unfortunately, it is not clear that only three levels (‘good’, ‘marginal’, ‘bad’) of acceptability exist, nor that three levels are adequate to explain perceived differences. Examples can be constructed which illustrate five or more levels of gradedness. Kluender (1992), for instance, offers the following acceptability ordering (from best to worst) of CNPC violations:

(21) a. This is the paper that we really need to find someone who understands. ≥
    b. This is the paper that we really need to find a linguist who understands. ≥
    c. This is the paper that we really need to find the linguist who understands. ≥
    d. This is the paper that we really need to find his advisor, who understands. ≥
    e. This is the paper that we really need to find John, who understands.

Further levels of grammaticality could be stipulated in the grammar, of course, but it is not clear that any finite number would actually be adequate. And, as Chomsky says, if languages do not employ ‘counters’ for the purpose of tracking violations, it becomes difficult to justify the existence of some large number of grammaticality levels.

Island constraints in the Minimalist Program emerge as a consequence of the Minimal Link Condition (Chomsky, 1995; Chomsky, 2000), which constitutes the only locality constraint in the grammar, hence replacing both the Wh-Island Constraint and Chomsky’s (1973) Superiority Condition, which stipulated that one wh-expression cannot be fronted over another structurally higher wh-expression:

(22) The Minimal Link Condition [Chomsky, 1995: 311]
    K attracts α only if there is no β, β closer to K than α, such that K attracts β.

However, as Sabel (2002a) notes, the MLC fails to predict acceptability differences among wh-island violations, e.g. object wh-movement out of a wh-island is better than adjunct movement or subject wh-movement (examples taken from Sabel (2002a)):

(23) a. What do you wonder [how John could fix ___]?
    b. How do you wonder [what John could fix ___]?
    c. Who did do you wonder [ how ___ could fix the car]?

Sabel concludes that these empirical oversights warrant an expansion of grammatical machinery, rather than a reevaluation of the generalizations about islands. The precise details are not
critical for the current discussion. Our interest here is in documenting the fact that a theoretical choice has once again been made not to reassess the fundamental data that must be accounted for by grammar. Accordingly, even some of the most recent renditions of generative syntactic theory appeal to grammatical constraints to explain the bulk of island effects.

The island exceptions already noted in English are further accompanied by a panoply of counterexamples from other languages. Although the proposed notions of Subjacency, Barriers, and other island constraints purportedly constitute innate universal principles of language, many languages such as Swedish (Allwood, 1976; Engdahl, 1982; Andersson, 1982), Danish (Erteschik-Shir, 1973), Icelandic (Maling, 1978), Norwegian (Taraldsen, 1982), Italian (Rizzi, 1982), French (Sportiche, 1981; Hirschbühler & Valois, 1992), Akan (Saah & Goodluck, 1995), Palauan (Georgopoulos, 1985, 1991), Malagasy (Sabel, 2002b), Chamorro (Chung, 1994), Bulgarian (Rudin, 1988), Greek (Alexopoulou & Keller, 2003), Yucatec Mayan (Elizabeth Norcliffe, p.c.), and doubtless many others, exhibit clear counterexamples to this prediction.\footnote{To be clear, these languages are not identical when it comes to which island constraints are clearly violable. Some of these languages seem to readily permit dependencies into wh-islands, while other island constructions in the same language are nevertheless degraded in acceptability.}

\begin{equation}
\begin{align*}
&\text{a.} & \text{Den dår gämla skräphögen kanner ja killen som köpte?} \\
& & \text{That old piece junk know I the guy who bought}.
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{b.} & \text{Vilken bok kunde ingen minas vem som skrivit?} \\
& & \text{Which book could no one remember who that had written?}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{[Swedish: Andersson, 1982]}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{Petta er lagið, sem enginn vissi hver samdi.} \\
&\text{This is song-DEF that no one knew who wrote.}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{‘This is the song that no one knew who wrote.’}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{[Icelandic: Maling, 1978]}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{Pion anarotihikes an tha apolisoune?} \\
&\text{who-ACC wondering-2SG whether/if will fire-3PL}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{‘Who did you wonder whether they will fire?’}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{[Greek: Alexopoulou & Keller, 2003]}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{Voilà la personne [que vous ne sauriez imaginer [avec quelle sauvagerie la police secrète a essayé de [faire parler ]]]}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{‘This is the person that you can’t imagine with what brutality the secret police tried to get to talk.’}
\end{align*}
\end{equation}

\begin{equation}
\begin{align*}
&\text{[French: Hirschbühler & Valois, 1992]}
\end{align*}
\end{equation}
But cross-linguistic counterevidence to a given island constraint has not generally been taken to be a refutation of the universality of that constraint as part of Universal Grammar. Perhaps the most common approach to such counterevidence is to assume that the apparent island violations really constitute a dependency between a base-generated (i.e. non-moved) phrase and a null or overt resumptive (Georgopoulos, 1985, 1991; Saah & Goodluck, 1995). This strategy is essentially what is proposed by Rizzi (1990) and Cinque (1990), whose ideas we discuss in the next section: apparent violations of movement principles are categorized as cases of non-movement – that is, anaphor-antecedent dependencies. Unfortunately, the reasoning here quickly becomes circular: the apparent violations do not involve movement, because this would contradict island constraints; and since they do not contradict island constraints, island constraints have not been counterexemplified. Effectively, as long as a case can be made for the existence of at least some resumptives in a language, there is a ready-made way of reanalyzing apparent island violations of this kind.

What stands out in the history of research on syntactic islands, in our view, is the entrenchment of the generalizations that were formed from early datasets. Despite extensive counterexamples or graded datasets, reassessments of the generalizations embodied in the grammatical constraints that were first proposed to deal with particular phenomena (e.g. the AOAC) have been rare. Instead, problematic examples have been labeled as ‘special’, ‘peripheral’ or ‘unique’, without providing independent motivation for these assessments. As we will show, there is a plausible alternative approach to CNPC and Wh-Island phenomena that makes sense of both the problematic counterexamples and the graded datasets.

### 2.2 Factors Affecting the Acceptability of Islands

Beyond the graded nature of acceptability judgments observed in the primary literature on islands, a number of other sources have observed manipulations that affect the ‘naturalness’ of island constructions. What is notable about these manipulations is that they do not change or eliminate the syntactic configurations that are supposed to be the basis of the island constraint in question. As we shall see, however, these observations have not been interpreted...
as evidence against configurationally defined syntactic islands; instead, they have spawned a
series of proposals that proliferate mechanisms within the theory of phrasal movement. These
complications of grammar effectively remove more and more data from the scope of island
constraints, allowing the latter to remain part of the grammar, in spite of considerable apparent
evidence to the contrary, applying to an ever diminishing subset of the extraction data.

Perhaps the most widely discussed factor that affects the acceptability of \textit{wh}-extractions
is the nature of the \textit{wh}-filler-phrase. Karttunen (1977) observed that while a bare \textit{wh}-word
fronted over a bare subject \textit{wh}-word is judged as unacceptable (ostensibly due to the Superi-
ority Condition), \textit{which}-\textit{N} phrases improve the acceptability:

\begin{enumerate}
  \item Which medication did which patient get? \geq
  \item What did who get?
\end{enumerate}

Maling and Zaenen (1982) noticed, too, that \textit{wh}-Island violations sound better with a \textit{which}-\textit{N’}
phrase compared to a bare \textit{wh}-item:

\begin{enumerate}
  \item Which article don’t you remember who wrote? \geq
  \item What don’t you know who wrote?
\end{enumerate}

Similar sorts of observations regarding properties of the filler-phrase soon followed in Pe-
theorized that \textit{wh}-phrases can be ‘D(iscourse)-linked’ and that this property voids the normal
constraints on Superiority (or some comparable constraint that blocks (30b), e.g. the Minimal
Link Condition). On a D-linking account, (30a) is better than (30b) because the \textit{wh}-phrases
are D-linked in the former, but not in the latter example. Unfortunately, a formal definition of
D-linking is absent from the literature, making it virtually impossible to evaluate any theory
based on this notion. The most common use of the term is intended to mean roughly that the
set of possible answers is pre-established or otherwise salient. For numerous reasons, how-
ever, such a definition is ultimately inadequate (Hofmeister, Jaeger, Sag, Arnon, & Snider,
2007). First, regarding decontextualized linguistic examples like those in (30), the set of pos-
sible answers is no more pre-established in (30a) than it is in (30b). Second, nothing stands
in the way of having a pre-established answer set associated with a bare \textit{wh}-item, as Pesetsky
himself acknowledges. Finally, there is no account of how or why general discourse properties
like salience (sometimes) interact with the application of grammatical constraints, leaving a
black box at the center of the D-linking analysis.

Rizzi (1990) documents similar effects for Italian \textit{wh}-phrases, as in (32), where the com-
plex interrogative phrase \textit{A quale dei tuoi} leads to higher acceptability than a minimally dif-
ferent question with \textit{a chi}:

\begin{enumerate}
  \item A chi non ti ricordi quanti soldi hai dato?
      To whom don’t you remember how much money you have give.
      ‘To whom don’t you remember how much money you gave?’
  \item A quale dei tuoi figli non ti ricordi quanti soldi hai dato?
      To which of your kids don’t you remember how much money you have give
      ‘To which one of your kids don’t you remember how much money you gave?’
\end{enumerate}
To account for such contrasts, Rizzi proposed a division between NPs (or DPs) that are ‘intrinsically referential’ and those that are inherently nonreferential, defining the former as those arguments which “refer to specific members of a set in the mind of the speaker or pre-established in discourse.” Cinque also adopted this distinction, in order to deal with extraction phenomena involving a range of quantifier phrases in Italian. Complex *wh*-phrases like those in (32b) were thus categorized as referential, while *wh*-phrases like *a chi* were labeled nonreferential.

Dividing arguments into these two classes is explanatory on the critical assumption that there is a fundamental difference between the movement constraints that govern the two classes of expression. Rizzi and Cinque claimed precisely this, arguing that the movement traces of referential arguments do not need to be antecedent-governed – that is, the relationship between a referential filler-phrase and its trace is not constrained by structural considerations. The traces of adjuncts and nonreferential arguments, however, must still be antecedent governed. Thus movement of phrases of the latter type must proceed in a fundamentally different way, i.e. via successive-cyclic movement. Nonreferential phrases like *who* in English and *a chi* in Italian, therefore, are treated as being sensitive to structural constraints on movement, unlike referential arguments.

Rizzi and Cinque justified this division by appeal to the relationship between referentiality and the notion of ‘bearing a referential index’. In particular, they assumed that referential arguments carry referential indices, enabling a binding relation to be made via coindexing (as in ‘anaphoric binding’). Since binding relations remain unconstrained by syntactic boundaries, ‘long movement’ (movement that crosses syntactic island boundaries) of referential arguments is licensed. On the further assumption that nonreferential phrases lack such indices, their relationship to their traces must be licensed by some distinct mechanism.

However, as Chung (1994) argues, this reasoning is at odds with standard wisdom about bound variable anaphora, since argument phrases falling into the nonreferential class can clearly antecede a bound variable pronoun. That is, the above analysis would require a separate indexing system to relate ‘referential’ elements to their traces, essentially undermining the argument from binding:

“. . . Rizzi’s referential indices cannot be identified with the indexing mechanism that is a mainstay of current approaches to anaphora (see Frampton 1991:39-42 for similar conclusions). But if that is so, then we are left wondering whether the use of indices in this theory amounts to more than a diacritic to distinguish the DPs that allow long movement from those that do not.” [Chung, 1994, p. 33]

Even if the indexing proposal was without these problems, it inevitably runs into the issue of why movement possibilities should vary with, or be determined by, specificity. As Chung (p. 39) asks:

“Why should long movement be legitimized in just those cases where the trace ranges over a sufficiently restricted set? To put the question differently, what is it about the ability to narrow down the domain of *wh*-quantification ‘enough’ that makes it possible for strict locality to be violated?”
Neither Pesetsky’s D-linking account nor the indexing analysis of Rizzi and Cinque clarify or motivate the relationship between properties of the filler-phrase and the acceptability of FGDs. While they make the valuable contribution of identifying an acceptability effect based on the nature of the extracted element, they see the effect in terms of a sharp bifurcation of elusive grammatical mechanisms. But the violability of island effects is not an all or nothing matter, and a more successful approach may become available if one views the matter in terms of the interaction of grammar and factors causing processing difficulty.

Consider one additional factor that affects the acceptability of island violations, but which has not been taken as counterevidence to island theory. As mentioned in the previous section, islands demonstrate a sensitivity to the definiteness and/or specificity of intervening constituents. In his 1973 Subjacency theory, Chomsky highlighted this peculiarity in his discussion of picture NPs, but did not discuss why the definiteness of an intervening NP should interact with the application of movement transformations. He does, however, make the following statement, based on the observation that judgments surrounding picture NP extractions are not uniform:

“A refinement of the condition (26) [the Specified Subject Condition] incorporating the feature [definite] as well as the property of lexical specification might be proposed to accommodate these judgments. Specified subjects in NPs are [+ definite]. If (26) is revised to include [+ definite] as well as specified subjects, then (31b) [Who did you see John’s picture of] will involve a double violation and Who did you see the pictures of only a single violation. This might account for the gradation of acceptability.”

As with the accounts of Pesetsky, Cinque, and Rizzi, such a proposal offers no way of understanding why the definiteness of intervening NPs should significantly alter the applicability of movement transformations.\(^5\) Again, our concern here is not so much with the particulars or the adequacy of these early accounts, but with the sustained practice of dismissing problematic examples as either inconsequential or as evidence of some secondary grammatical mechanism.

It is interesting to note that the same sensitivity evidenced by English FGDs has also been documented in Swedish. While Swedish demonstrates a far greater allowance of island-violating FGDs, CNPC violations are appreciably worse when an intervening NP is definite:

(33) a. Johan känner jag ingen som tycker om.
    Johan know I no one that likes
    ‘Johan, I know no one that likes.’

b. Johan känner jag en flicka som tycker om.
    Johan know I a girl that likes
    ‘Johan, I know a girl that likes.’

\(^5\)Manzini (1992) attempts to unify the effects of both tense and definiteness on islands. Her suggestion is that both tense and definiteness block movement out of islands, because they interpose a conflicting case address. However, despite the evidence that definite interveners and tensed clauses make island-violating dependencies worse, we note that finiteness and definiteness do not seem to be categorical predictors of grammaticality, even in island contexts, as long as other processing burdens are removed.
Relative clauses within nonspecific or generic NPs tend to be easier to extract out of than relatives that are part of definite NPs; however, as Andersson (1982) points out, this is merely a tendency and extractions out of relative clauses with definite heads are indeed possible.

This sensitivity to the referential properties of intervening constituents in both Swedish and English intriguingly matches the contrast observable in center-embeddings, where decreasing the specificity of embedded subjects improves the examples (although the construction remains difficult to process):

(34) a. The host [someone [I knew ___] brought ___] left. ≥
    b. The host [the boy [the girl [I knew ___] brought] left.

Here, as in the picture NPs and the Swedish CNPC violations, nonspecific or indefinite intereners improve comprehensibility and acceptability. In fact, the sensitivity to the properties of intervening constituents does not appear to be limited to nominal arguments. The frequency, specificity or ‘semantic richness’, verb class, and argument selection properties of verbs on the extraction path also affects acceptability judgments. In Swedish, for instance, extraction out of complex NPs (putative CNPC violations) is preferable when the matrix verb is stative:

(35) a. De blommorna känner jag en man som säljer.
    Those flowers know I a man who sells
    ‘Those flowers, I know a man who sells (them).’
    b. De blommorna talar jag med en man som säljer.
       Those flowers talk I with a man who sells
       ‘Those flowers, I talk with a man who sells (them)’

In English as in Swedish, the acceptability of dependencies varies with properties of intervening verbs. It seems, for example, that stative verbs are preferable to activity verbs [(36)] and that communication verbs are preferred when they do not specify the manner of speaking [(37); Erteschik-Shir, 1973]:

(36) a. Which book did you see pictures of?
    b. Which book did you destroy pictures of?

(37) a. Who did you say that John believes you saw?
    b. Who did you lisp that John believes you saw?

Additionally, the acceptability of sentences containing FGDS looks to be generally caught up with features tied to intervening verbs – and not merely a sensitivity that is observable in island-violating constructions. Hence, Deane (1991) categorizes the first two examples below as acceptable, but the latter two are marked or worse in comparison:

6The use of the indefinite NP and personal pronoun also creates a more distinctive sequence of NPs than three definite NPs, one after another. The effects of similarity are discussed in the next section.
Deane notes that the interpretation of the latter two involves less plausible or likely scenarios (criticizing or buying votes) compared to the first two examples (obtaining or finding votes).

Cumulatively, the acceptability of FGD sentences varies in seemingly systematic ways with the nature of material that intervenes along the filler-gap path. This fact is irreconcilable with the syntactic proposals outlined in the previous section, since they uniformly contain no mechanism for evaluating lexical properties and adjusting the possibilities for movement according to these properties. As illustrated by Chomsky’s proposal, island constraints could be embellished with additional featural specifications, but these elaborations are fundamentally ad hoc – there is no accompanying theory of why these features should impact dependency formation.

The literature on islands has also exhibited a steadily increasing acknowledgment that not all island phenomena can be explained on structural grounds alone. Alongside the attempts to preserve the core syntactic generalizations, semantic and pragmatic explanations for Subjacency effects have been proposed as well, especially for the case of weak islands like tenseless \(wh\)-islands, factive islands, and negative islands (Erteschik-Shir & Lappin, 1979; Kroch, 1998; Comorovski, 1989; Rizzi, 2000; Szabolcsi & Zwarts, 1993; Szabolcsi, 2006; Oshima, 2007; Truswell, 2007). In general, these semantic and pragmatic accounts operate from the starting point that not all islands are equally bad, and that syntactic constraints cannot or should not explain these differences. For instance, extraction out of a tenseless \(wh\)-clause is considered to be mildly degraded or marginally acceptable, compared to extraction out of a tensed \(wh\)-clause:

(39) This is a topic which John wondered whether to talk about. ≥
    This is a topic which John wondered whether she talked about.

In fact, the defining property of so-called weak islands is that they block movement of some but not all phrases. For example, tenseless \(wh\)-islands have been claimed to allow argument, but not adjunct extraction (Huang, 1982). Consequently, they are contrasted with strong islands which purportedly block movement of all phrase types.\(^7\)

Kroch (1998) also rejects a ‘pure’ syntactic explanation for island phenomena, arguing that some island effects, like the ban on moving adjunct phrases out of \(wh\)-islands and negative islands, owe their status to pragmatic considerations, rather than purely syntactic constraints. Others have come to a similar conclusion with respect to still other island effects: putative island violations owe their status to pragmatically aberrant, or even paradoxical interpretations. That said, these semantic and pragmatic accounts do not dismiss the validity of syntactic island constraints in general. Instead, they implicitly aim to preserve the essentials of the syntactic

\(^7\)However, as noted by Szabolcsi (2006), this “dichotomy is not particularly straightforward; moreover, the borderline between strong and weak islands is not very firm, there being a number of intriguing empirical parallels between the two.”
theories by claiming that seemingly problematic data should be explained in terms of interacting nonsyntactic factors. As a consequence, the standard syntactic island constraints are left in place. The semantic and pragmatic factors function so as to ‘explain away’ problems that can then be safely ignored by purely syntactic theories of islands.

In essence, our approach here is the same as the semantic and pragmatic accounts. That is, we are reinterpreting some of the island data and suggesting that syntactic principles cannot account for the observed variations in acceptability. In this sense, our analysis is not as radical as it first appears – it leaves intact the possibility that some island phenomena originate from syntactic constraints. The major difference between our analysis and the semantic-pragmatic theories mentioned above is that we believe that once the processing burdens are properly understood (and explained partly in terms of semantic and pragmatic factors), there remains little work to be done by purely syntactic island constraints.

In sum, the island data we have surveyed is characterized by graded and variable judgments, as well as exceptions within and across languages. Historically, the major analytic proposals of this data have attempted to present a unified syntactic theory, but these have been repeatedly met with clear counterexamples and problematic subtleties of graded acceptability. In turn, this has led to a panoply of accompanying proposals aiming to preserve the core principles of the syntactic theory, while explaining away the ‘peripheral’ data that counterexemplify them. While on some level there is consensus within the field that the full set of island data cannot be dealt with in purely syntactic terms, there is still a strong impetus to approach as much of the data as possible in purely syntactic terms.

3 Processing Islands

Despite the early recognition of the role that processing pressures play in determinations of acceptability, the syntactic literature has by and large avoided invoking performance-related factors to account for either categorically different perceptions of grammaticality, or even graded perceptions of grammaticality. This situation is especially puzzling in the case of island phenomena because these are characterized by so many features that have been known independently to contribute to processing difficulty. This section catalogs some of the most significant factors that are likely to influence the processing of island constructions (but it is not meant to be exhaustive). Of course, the structure of each island construction makes certain factors more or less relevant. To understand the argument that processing costs may be able to explain certain island phenomena, it is crucial to understand the widely accepted principle that the cognitive systems underlying language avail themselves of limited resources. To expend cognitive resources in one part of a sentence is thus likely to reduce the availability of resources to process subsequent parts of that sentence.

The defining feature of all island violations is, of course, that they contain filler-gap dependencies. It has been clear for some time that a sentence with a FGD incurs a relatively high degree of processing difficulty, compared to a minimally different sentence without the dependency (Wanner & Maratsos, 1978; King & Just, 1991; Kluender & Kutas, 1993a; Hawkins, 1999):

“The filler-gap dependencies are difficult structures to process... Identifying the gap...
is not easy. It is an empty element with no surface manifestation and its presence must be inferred from its immediate environment. At the same time, the filler must be held in working memory, and all other material on the path from filler to gap must be processed simultaneously, and the gap must be correctly identified and filled.” [Hawkins, 1999, p. 246-7]

All sentences with FGDs are therefore more costly to process. Note that this additional processing cost is not without purpose: it pays for a noncanonical information structure or it otherwise realizes some discourse goal that a minimally different construction lacking such an FGD would not be able to accomplish. In other words, the special discourse functions that are associated with FGDs come with a processing cost.

Psycholinguistic research has confirmed not only that processing load increases inside a dependency, but also that processing difficulty generally increases as the dependency gets longer (Gibson, 1998; Gibson & Pearlmutter, 1998; Gibson, 2000; Hawkins, 1999; Grodner & Gibson, 2005; Fiebach, Schlesewsky, & Friederici, 2001). The standard explanation for this effect is that the activation of a mental representation decays over time, making it progressively harder to efficiently reaccess the memory item:

“...the greater the distance between an incoming word and the head or dependent to which it attaches, the greater the integration cost; and ... the longer a predicted category must be kept in memory before being encountered, the greater is the cost for maintaining that prediction.” [Gibson & Pearlmutter, 1998, p. 265]

This leads to a preference for local attachment in ambiguous structures, and it also buttresses the claim that there is a preference associated with positing a gap as soon as possible (Stowe, 1986; Frazier, 1987). But this memory-based constraint on dependencies also predicts processing differences for unambiguous dependencies where the length of the dependency varies.

Accordingly, prior research has established the existence of processing differences between nested and nonnested structures with identical meanings:

(40) a. [The scientist collaborated with the professor [who had advised the student [who copied the article]]].
   b. [The student [who the professor [who the scientist collaborated with] had advised] copied the article].

In the nonnested sentence [(40a)], the distance between dependents is minimal, making the sentence relatively easy to process. But the nesting or center-embedding process evidenced in (40b) increases the distance between the dependents, which raises the retrieval and integration costs significantly.

Many island-constructions, including CNPC and WH-Island violations, are of course characterized by relatively long filler-gap dependencies. This feature, therefore, makes them particularly susceptible to the influence of other compounding comprehension difficulties. To

---

8There are, however, some important exceptions to this principle. Generally, these exceptions involve material interceding between filler and gap that increases the predictability of the subcategorizing head or else improves the retrievability of the dependent phrase (see, for instance, Vasishth & Lewis, 2006).
be clear, dependency length alone and the associated processing difficulty cannot account for the unacceptability of islands, as sentences with comparably long dependencies are generally considered acceptable tokens of English. This onus of maintaining a dependency over relatively long periods, however, is frequently accompanied by other linguistic factors that further heighten processing difficulty.

It is important to understand, moreover, that increasing dependency distance in unquestionably grammatical contexts does reduce acceptability, in accord with the assumption that increased processing difficulty lowers acceptability. Arnon et al. (2005), for instance, verified that subject *wh*-interrogatives were judged more acceptable than object *wh*-interrogatives:

(41) a. Which man ___ saw the girl in the bar on California Avenue? ≥
    b. Which man did the girl in the bar on California Avenue see ___?

In the subject *wh*-question, the *wh*-phrase can essentially be integrated and assigned a semantic role at the very next word (*saw*). By contrast, the *wh*-phrase in the object question must be stored and maintained in memory while three additional discourse referents are identified, categorized, and integrated. Hence, even in quite commonplace FGDs, the distance between the filler and gap can have a significant bearing on both measurable processing difficulty and judgments of acceptability, as obtained in controlled experimental circumstances.

We have already discussed the fact that the specificity or referentiality of phrases appearing between a filler-phrase and its gap influence perceptions of acceptability. Chomsky and others have noted that properties of intervening constituents appear to interact with the possibilities for movement, but there has been no principled account of why such interactions should occur. From a psycholinguistic perspective, the impacts of specificity and/or referentiality are part and parcel of a more general pattern in dependency processing: intervening constituents whose processing consumes more resources reduce the resources available to link the filler with its gap. In other words, FGDs are processed more efficiently when the material between the filler and gap is itself easier to process.

Advocating this line of thought, Deane (1991) suggests that the ability to form a dependency is contingent upon the ability to attend to both the extracted element and the retrieval site simultaneously – in other words, the ability to retrieve an extracted element (or have it in some sort of attentional focus, in his terms) and integrate it at the correct gap site. These attentional demands are impaired by distractions that occur along the filler-gap path:

“Now, in a sentence involving extraction, we have hypothesized that the extraction site and the extracted NP must command attention. Of course, they would elicit attention most readily in the absence of competing elements (distractors). It would follow that the rest of the sentence should be relegated to the background, that is, presupposed or at least given . . . Failures of extraction come when these conditions are not simultaneously satisfied.” [Deane, 1991, p. 36]

Along these lines, the effects tied to the referentiality and specificity of intervening NPs are also illuminated by processing studies. Among other results, Warren and Gibson (2002) provide reading time evidence that definite NPs and proper names that intervene along a filler-gap path cause slower reading times at the retrieval site than do intervening personal pronouns:
The consultant who { Donald Trump, the chairman } avoided at the party.

Warren & Gibson attribute this cost difference to mental accessibility of the nominal referents, noting that first and second person pronouns are old referents in discourse:

“...the effort spent attempting to access a referent before quitting and instantiating a new one may be less for NP types whose referents must be highly activated than for NP types whose referents are usually less activated.” [Warren & Gibson, 2005, p. 754]

Ariel (1990) and others have developed scales of ‘accessibility’ with informationally light elements like pronouns on the high end of the scale, and informationally heavy expressions like definite descriptions much lower on the scale. The guiding intuition here is that the less salient (and therefore harder to retrieve from memory) a referent is, the more instructions a listener needs in order to identify the appropriate referent. Hence, when it is already clear what the intended referent of an anaphor is, short and nonspecific forms are licensed – few instructions have to be parsed and interpreted, which means less processing work. In contrast, when the referent is less salient, more instructions are needed to ensure successful comprehension, e.g. a definite description. And in order to parse and comprehend more instructions, more resources must be expended.

Now add to this picture the previously mentioned principle that there are limited cognitive resources available for linguistic processing, and expending those resources in between a filler-phrase and its gap increases the overall processing difficulty of the sentence (Warren & Gibson, 2005). On this view, the acceptability variation evident in extraction out of picture NPs, complex NPs, and other island constructs is expected, considering the findings that relate processing difficulty to the accessibility or complexity of the intervening constituents. Certain types of NPs engender more processing difficulty than others because they contain more information or instructions for identifying and/or situating referents, e.g. someone vs. the girl. Definiteness effects, as in the picture NPs discussed by Chomsky and others or the Swedish CNPC examples, can also be understood as derived from processing factors. Since definites overwhelmingly refer to old or established referents in the discourse, such NPs may automatically trigger a search for a referent. In the decontextualized examples that linguists routinely construct, where there is no contextually established antecedent, such searches will ultimately fail and a new mental representation will be created to accommodate the presuppositional demands of the sentence. In real conversation, where context does provide an antecedent, the mental search will still have to be conducted, though overall the processing task is easier. An indefinite, however, would not trigger a search in either situation, and a new mental representation would immediately be created, thus bypassing a potentially costly processing task.9

One of the fundamental observations behind syntactic accounts of islands is that certain syntactic configurations impose barriers or boundaries to movement. For instance, \textit{w/h}-clauses

---

9 Other discourse considerations also factor into the cost of definites versus indefinites (and other phrasal types). For instance, old or given information is more likely to appear early in the sentence, while new information typically shows up later in the sentence.
limit extraction possibilities, purportedly because the grammar specifies restrictions that
ensure this. Experimental findings from Kluender & Kutas (1993), however, raise the possibility
that boundary-based effects ultimately stem from processing considerations and not spec-
ifications in the syntax. They report that, even in yes-no questions where no filler is carried
across the clause boundary, different complementizers elicit significantly different neurophys-
iological responses, as well as acceptability judgments. The complementizer that produces
the highest rating of acceptability, while the bare wh-word who garners the lowest rating:

(43) a. Has she forgotten [that he dragged her to a movie on Christmas Eve]?
    b. Has she forgotten [if he dragged her to a movie on Christmas Eve]?
    c. Has she forgotten [who he dragged ___ to a movie on Christmas Eve]?

These findings suggest that there is indeed something special about embedded interrogatives,
as in (43a) and (43b). They elicit neurological responses compatible with an interpretation of
greater processing difficulty, and judgments that reflect degraded acceptability. More gener-
ally, different types of syntactic boundary appear to have different cognitive costs, independent
of whether or not a FGD crosses that boundary.

The difficulty of processing FGDs is contingent not only on the length of the dependency
and the number and kind of intervening constituents, but also properties of the filler itself.
Hofmeister (2007) provides extensive evidence that the quantity of information encoded in a
filler affects filler-gap processing in a wide variety of sentential contexts. In particular, syntac-
tically and semantically more complex filler-phrases, while initially requiring more resources
to process, are shown (via reading time studies) to facilitate processing at retrieval points.
For example, definite NPs like the one in (44b), which is syntactically and semantically more
complex than the definite in (44a), were found to produce faster reading times beginning at
the subcategorizing verb (encouraged):

(44) a. The diplomat contacted the dictator who the activist looking for more contribu-
tions encouraged to preserve natural habitats and resources.
    b. The diplomat contacted the ruthless military dictator who the activist looking for
more contributions encouraged to preserve natural habitats and resources.

The same complexity effects appear in clefts and relativizations, as well as in non-island wh-
dependencies. Moreover, these complexity-based effects have been found to occur with wh-
phrases and definites, as well as indefinite NPs. To account for this pattern, Hofmeister (2007)
argues that linguistic expressions that contain more syntactic and semantic information fa-
cilitate later retrieval of the corresponding linguistic representations from memory. In other
words, more elaborate descriptions present better targets for recall than vague descriptions.
Compared to a very specific and detailed description, an uninformative or vague description
will not be distinctive in memory and is more likely to overlap in features with other represen-
tations in memory. This memory-based account thus offers a potential means for explaining
why semantically richer, more complex filler phrases promote increased acceptability, as ob-
erved by Pesetsky, Rizzi, Cinque, and others.

Another common problem with many of the examples of island violations cited in the
syntactic literature is that they often lead to an initial misparsing of the sentence. That is, given
a preference for associating a filler with a head as early as possible, i.e. an active-filler strategy, the parser will try to integrate a filler-phrase stored in memory at the first possible opportunity. Thus, given the partial string *Who did you see* . . . , the *wh*-item *who* could reasonably act as the object of the verb *see*, so the parser is likely to attempt integration at this site. If more material follows, disconfirming this initial analysis, then the original parse has to be revisited, creating processing difficulty.

As noted earlier, this list of processing factors is hardly exhaustive. Sentence processing research has also identified critical roles for frequency or predictability in determinations of processing difficulty – less frequent or unpredictable constituents are more difficult to process (Hale, 2001; Jurafsky, 2003; Levy, 2005). Other sentence processing factors that are potentially relevant for the processing of islands include similarity-based interference (Lewis, 1996; Gordon, Hendrick, & Levine, 2002), collocation frequency, contextualization, as well as complexity variation due to semantic type (e.g. question versus proposition).

Cumulatively, islands are wrapped up with numerous challenges for sentence processing. From island to island, these cognitive challenges undoubtedly have different weights and effects. Indeed, certain island effects may be quite independent of processing difficulty. Hence, although some island-inducing constraints may be part of grammar, the evidence is compelling that many island constructions have characteristics which have been independently identified as contributing to processing difficulty.

An analysis incorporating these processing facts can side-step the need to posit ‘special’ rules or a difference between normal and ‘special’ speakers. Our central point with respect to processing costs in islands is that there are *multiple* processing burdens that can interact to create unacceptability as the biproduct of multitask interference. Moreover, this interaction is potentially nonlinear in nature. In other words, factors that consume the same or conflicting cognitive resources can interact in a highly nonlinear (or ‘superadditive’) fashion. The combined effects of multiple processing hinderances do not have to add up to their individual effects if they tap the same pool of cognitive resources. Definiteness by itself, therefore, may contribute to an overall perception of reduced acceptability, but definiteness in conjunction with one or more further resource-consumptive factors may constitute a serious obstacle to sentence processing, and may hence dramatically reduce acceptability.

Given these strong indications that the variation in islands may reflect processing preferences, we now turn our attention to a series of experimental investigations that explore the role of processing in island effects. These studies are intended to evaluate the extent to which acceptability differences are tied to processing differences. A theory that suggests that processing differences underlie island effects must verify that the processing differences actually exist. If controlled experiments find that acceptability differences are not accompanied by processing differences in islands, then the evidence would obviously not support a processing-based explanation. Similarly, if the results turn out to be the opposite of what is predicted by our knowledge of cognitive constraints, then the processing hypothesis would be similarly challenged. On the other hand, if processing differences within islands are systematically matched with contrasts in acceptability judgments, then this would be compatible with the hypothesis that limitations on cognitive resources are responsible for island effects. To be clear, there are other possible interpretations of such findings, e.g. that grammatical differences underlie the processing differences. In section 8, however, we present a number of arguments that favor
the conclusion that cognitive constraints are responsible for the island phenomena we have investigated.

4 Methodology

In the following sections, we present the results of three experiments that utilize the self-paced reading methodology. In these comprehension experiments, subjects read sentences at their own pace on a computer screen (Just, Carpenter, & Woolley, 1982). Participants are presented with a screen of dashes separated by spaces, representing the words for that experimental item. With each press of a predefined key, a new word appears on the screen and the previous word disappears. An example of a sequence of such states is given below:

(45) —– call — —— ———— — ———– — — ———
——— to ——— ———— — ——— — — ———
——— —confirm ——— ——— — — ———

The amount of time between each key-press is automatically recorded and tagged with relevant information. Longer reading times at a particular word or region are interpreted as an indication of processing difficulty.

Between six and eight practice examples preceded the real experimental items in order to acquaint participants with this manner of reading and to reduce the magnitude of order effects on items presented early in the experiment. During the experiments proper, each participant saw each item in exactly one condition (Latin-square design). Blocking of items into lists and randomization within lists was automatically managed by the reading-time software, LINGER, v. 2.94 (Rohde, 2003).

After reading a comprehension question, subjects responded either by selecting the correct answer from a set of possible choices or by responding ‘yes’ or ‘no’ in the case of polar interrogatives. Response times for answering the question, as well as question-answer accuracy, were also studied. If the mean question-answer accuracy for a subject was below 67%, then the entire data set for that individual was dropped from the analysis. If a subject’s global reading time average differed from the sample’s global average by more than 2.5 standard deviations, that subject’s results were also excluded. Since reading times have a naturally imposed minimum time, subjects with global reading averages under 200 ms. were excluded in addition, even though this reading time might not have been excluded, if only standard deviations were considered. Due to poor question-answer accuracy for some items in Experiment I (7 out of 252 cells, or 2.77% of the item analysis data set), the missing data for those cells were replaced with the linear trend for that point using the Replace Missing Values command in SPSS 15.0. This method of data imputation inserts the predicted values of empty data cells by regressing the existing series on an index variable scaled 1 to \( n \).

Residual reading times were derived for each word on the basis of a linear regression equation that computed reading time for each individual as a function of word length. A subject’s residual reading time for a given word of length \( n \) expresses how fast that word was read compared to that subject’s average reading time for other words of length \( n \). So a negative residual reading time means a comparatively fast reading time. The practice of using residual
reading times reduces variability due to word-based differences (Ferreira & Clifton, 1986). Statistical outliers were removed by deleting residual reading times that were more than 2.5 standard deviations from the mean at each word region. This process affected less 3% of the total data in Experiment I, less than 2% in Experiment II, and less than 4% in Experiment III. Moreover, only reading times from correctly answered stimuli are considered here.

All experimental data were analyzed with repeated measures ANOVAs. In experiments where more than two treatment or condition levels are compared with one another, the results of planned t-tests are also reported. As repeated hypothesis testing with the same sample increases the likelihood of false positives (Type I error), the family-wise error rate must be corrected in such situations. Within the text itself, we report only the uncorrected p-values, but since more than three treatment levels are never compared with one another, the Bonferroni corrected level of significance can be safely estimated at .0167.

5 Experiment I: Complex Noun Phrase Constructions

The first experiment looks at the role of two independent processing factors in the processing of CNPC violations: the complexity of the displaced wh-phrase and the type of island-forming NP. Both factors have been identified as determinants in the overall acceptability of CNPC violations, as outlined above. In particular, increasing the complexity of the filler-phrase appears to raise acceptability judgments, while extraction out of a complex definite NP intuitively seems harder than extraction out of an indefinite.

If these acceptability effects are connected to processing effort, then manipulating these factors should produce observable processing differences. More specifically, relatively complex fillers like which-N phrases should lead to facilitated processing, as compared to simple fillers like bare wh-words. Similarly, if the observed effects of definiteness are tied to processing, then definite islands should be accompanied by greater processing difficulty than indefinite ones.

5.1 Materials & Participants

The stimuli in this experiment consisted of 36 embedded CNPC violations, which varied with respect to the information content of the filler-phrase and the nature of the island-forming NP. For each item, each participant saw one of seven experimental conditions (2 x 3 + 1). Across conditions, the fronted wh-phrase was either a bare wh-item (=BARE condition), such as who or what, or a comparatively more complex which-N phrase (=WHICH condition). Precisely half of the items presented animate wh-phrases, while the other half contained inanimate wh-phrases. The other factor considered in this experiment was the effect of NP type on subsequent sentence processing. Subjects read one of three kinds of island-forming NPs: a definite NP (DEF), an indefinite plural (PL), or an indefinite singular (INDEF). Additionally, a baseline for each item was included that lacked the island-forming NP. (46) shows a sample experimental item with all seven conditions:

(46)  BARE-DEF: I saw who Emma doubted the report that we had captured in the nationwide FBI manhunt.
BARE-PL: I saw **who** Emma doubted reports that we had captured in the nationwide FBI manhunt.

BARE-INDEF: I saw **who** Emma doubted a report that we had captured in the nationwide FBI manhunt.

WHICH-DEF: I saw **which convict** Emma doubted the report that we had captured in the nationwide FBI manhunt.

WHICH-PL: I saw **which convict** Emma doubted reports that we had captured in the nationwide FBI manhunt.

WHICH-INDEF: I saw **which convict** Emma doubted a report that we had captured in the nationwide FBI manhunt.

BASELINE: I saw **which convict** Emma doubted that we had captured in the nationwide FBI manhunt.

Eighty filler items of comparable sentence length accompanied these experimental items. Comprehension questions followed every experimental stimulus. These questions were polar interrogatives that probed subjects’ understanding of the sentence, e.g. **Was Emma sceptical that we had captured someone?** Participants were provided negative feedback if they answered a question incorrectly.

Thirty-one Stanford University undergraduates, all of whom were native English speakers, participated in this study; however, the results of six subjects were dropped, due to question-answer accuracies below 67% or reading times 2.5 standard deviations from the sample mean. Participants were paid $15 to complete this experiment and an unrelated off-line survey.

### 5.2 Results

We evaluate here the reading times for the word regions following the *wh*-phrases to word regions inside the embedded complement clause. Of particular interest are the reading times inside the complement clause, since this region constitutes a proposed island for extraction in the island-violating conditions.

Immediately after the *wh*-phrases, the complexity of the *wh*-phrase does not have a significant impact on reading times (NP type is, of course, irrelevant as the NPs have not been reached yet). On the second word after the *wh*-phrase, there is a main effect for *wh*-phrase complexity, such that the **less complex** BARE conditions lead to significantly faster reading times ($F_1(1,24) = 5.903, p = .023, F_2(1,35) = 4.220, p = .047$).

Starting at the complementizer *that*, however, the WHICH conditions generate significantly faster reading times than the BARE conditions ($F_1(1,24) = 7.450, p = .012, F_2(1,35) = 14.722, p < .001$). In all three NP-type conditions, as shown in Figure 1, the WHICH version is processed faster than the corresponding BARE version. This processing advantage for the WHICH conditions extends beyond the complementizer to include the subsequent pronominal subject, the embedded auxiliary and verb (see Figure 2), as well as the regions after the verb where the presence of the gap is confirmed. In other words, from the complementizer until several words after the subcategorizer, the WHICH conditions are read faster than the corresponding BARE conditions. Combining the regions inside the embedded clause, there is a highly significant main effect of *wh*-phrase type ($F_1(1,24) = 17.071, p < .001; F_2(1,35) = 21.195, p < .001$).
These results are summarized in Table 1.

[FIGURE 1 ABOUT HERE]

[FIGURE 2 ABOUT HERE]

Also of importance is the fact that, throughout the most embedded clause, the average reading time for the WHICH conditions does not significantly differ from that of the syntactically simpler baseline. Figures 1 and 2 reflect this point: reading times for the baseline are comparable to the average reading time of the WHICH conditions. This stands in stark contrast to the BARE condition, which remains consistently slower than the baseline throughout the embedded clause.

At only two word regions, however, are there main effects of NP type: at the complementizer and at the embedded verb. At the complementizer, the singular indefinite conditions are read faster than both the definite and plural indefinite conditions, which creates a main effect of NP type (F1(1,24) = 7.397, p < .01; F2(1,35) = 11.580, p < .001). Since the difference between the indefinites and the other two NP types is far greater in the BARE conditions, there is also an interaction of wh-phrase type and island NP type, but this interaction is only significant by items (F1(1,23) = 1.648, p = .214; F2(1,34) = 6.972, p < .01). At the embedded verb, there is also a main effect of NP type (F1(1,24) = 5.173, p = .014, F2(1,35) = 3.746, p = .034); in this case, the effect is largely driven by the fact that the BARE-DEF condition elicits significantly slower reading times than every other condition. No other comparisons at this region are statistically significant. Additionally, there are no other significant interactions at any word region.

In sum, the reading results identify a brief slowdown after processing the complex wh-phrases, but beginning with the complementizer, this disadvantage reverses and becomes a highly significant processing advantage. This advantage persists until several words after the subcategorizing verb. Effects due to the type of the island-forming NP are much more localized (and relatively weaker) and appear only at the complementizer and verb. Nevertheless, these effects do indicate that definite NPs lead to slower reading than indefinite NPs.

Reaction times to the comprehension questions show no effect of either complexity or NP type (BARE: 2555.74, SE = 65.87; BASELINE: 2280.27, SE = 95.42; WHICH: 2544.01, SE = 61.43). The BASELINE, however, produces faster reaction times than either island condition. This effect is significant by subjects, but only marginal by items (BASELINE-BARE: t1(24) = -3.713, p = .001; t2(35) = 1.609, p = .117; BASELINE-WHICH: t1(24) = -3.179, p = .004; t2(35) = 2.425, p = .021). [COMPREHENSION ACCURACIES]

[TABLE 1 ABOUT HERE]

5.3 Discussion

Reading times in this study show a strong influence of the complexity of the filler-phrase. While reading times immediately after the dislocated wh-phrase reflect slowed reading after a complex which- N phrase, reading times throughout the most embedded clause are significantly faster in the WHICH condition. In fact, reading times within the most deeply embedded
clause in the WHICH condition are not significantly different from reading times in the non-island baseline condition. In contrast, despite the BARE and BASELINE conditions both having bare wh-fillers, reading times in the BARE condition are significantly slower in the embedded clause. Hence, the which-N phrase largely offsets the difficulty introduced by the added complexity of the CNPC violation.

Although the effect of filler complexity is powerful, the definiteness of the island-forming NP was shown to have a weaker effect. Nevertheless, at two word regions, there is an observable effect of NP type. Moreover, the reading time differences at these sites line up with the prediction that singular indefinites lead to faster processing than definites. The results for the plural conditions, however, are harder to interpret, because this condition sometimes patterns with the definite NPs, and sometimes with the singular indefinites.

To complement the self-paced reading experiment and to understand the relationship between these comprehension results and judgments of acceptability, we also conducted a controlled acceptability study using identical stimuli, including the same fillers. As with the self-paced reading task, items were randomized and distributed across lists such that each subject saw only one condition per item. Sixteen subjects (none of whom had participated in the previous reading time study) were asked to rate the sentences for naturalness on a scale of 1 to 7 and were specifically instructed not to rate the sentences according to prescriptive grammar rules. The subjects in this study were given course credit for their participation. A mean acceptability rating was derived for each subject, based on all items and all fillers. The scores for individual items were divided by this overall mean to produce a judgment ratio that takes into account individual variation in usage of the scale.

As shown in Figure 3, the acceptability judgments yield the same main effect of wh-phrase complexity as the reading time study ($F_1(1,15) = 32.658, p < .0001; F_2(1,35) = 22.984, p < .0001$). Just as which-N phrases improve processing of the CNPC violations, they also raise ratings of acceptability ($t_1(15) = 5.668, p < .0001; t_2(35) = 5.033, p < .0001$). The baseline in the acceptability study receives significantly higher ratings than the WHICH conditions ($p < .001$). This contrasts with the reading time study, where the baseline did not consistently produce significantly faster reading times inside the critical embedded clause.$^{10}$ This contrast suggests that acceptability surveys can be sensitive to factors that do not noticeably impact on-line performance measures such as reading times. In particular, reading times may not reflect the effects of semantic, pragmatic, and/or discourse processing that extend beyond the first-pass reading of the sentence. Indeed, we find that the comprehension question response times for the baseline case in datasets like (46) are about 270 milliseconds faster than those of the WHICH-conditions. Thus, in comparing reading times and acceptability judgments, some differences are definitely expected, on the assumption that acceptability judgments constitute

---

$^{10}$ There is a very marginal effect of NP type in the acceptability results. In particular, as pictured in Figure 4, plural indefinites in the WHICH condition produced slightly better sounding CNPC violations, as compared to singular indefinites and definite NPs. All other comparisons of NP type were found to be nonsignificant.
the end-result (or involve the input) of a number of cognitive processes, including sentence processing effort. In sum, the acceptability ratings, in accord with the results of the comprehension study, reflect a preference for constructions with extracted \textit{which-\textbar N} phrases over bare \textit{wh}-words.

6  Experiment II: Wh-Islands

In this second experiment, we consider whether \textit{wh}-islands demonstrate the same sensitivity to filler complexity as reflected in the study of CNPC violations. As in the case of CNPCs, \textit{wh}-island violations are typically judged to be better when the displaced element is a \textit{which-\textbar N} phrase, instead of a bare \textit{wh}-item. To understand whether these differences are related to processing differences, we test the effect of such manipulations on comprehension. As in the previous experiment, a processing account of the previously noted acceptability differences predicts accompanying processing contrasts. Specifically, \textit{which-\textbar N} phrases, as opposed to bare \textit{wh}-words are predicted to lead to more efficient processing in island violations.

6.1  Materials & Participants

All experimental items contained \textit{wh}-islands, presented as main clause interrogatives, as in (47). Subjects were initially presented with a declarative background sentence. After finishing this sentence, a basic comprehension question was presented that asked about the identity of the embedded object NP. The targets of interest were the comprehension questions themselves; the initial context sentences merely justified the presence of these questions which would be unnatural without any background.

Stimuli varied in terms of whether the sentence-initial \textit{wh}-phrase was a bare \textit{wh}-item \textit{who} (\textsc{bare}) or a more complex \textit{which}-phrase (\textsc{which}). An additional condition was included to serve as a baseline against which the results could be compared. This non-island baseline condition always began with \textit{who}, but instead of a \textit{wh}-phrase in the complementizer position, these baselines items had the complementizer \textit{that}:

(47)  Albert learned that the managers dismissed the employee with poor sales after the annual performance review.

\textsc{bare}: Who did Albert learn whether they dismissed after the annual performance review?
\textsc{which}: Which employee did Albert learn whether they dismissed after the annual performance review?
\textsc{baseline}: Who did Albert learn that they dismissed after the annual performance review?

Twenty-four experimental items and forty-eight fillers constituted the materials for this study. Twelve of the fillers were also \textit{wh}-islands, so that half of the overall items were \textit{wh}-islands and the other half were not syntactic islands. The additional twelve \textit{wh}-islands were included to
mask some of the more salient properties of the true experimental items. For instance, while the twenty-four experimental items always asked about the object of the embedded verb, some fillers asked about a prepositional object, e.g. *Which legislation did Andrew report whether they reached a compromise over?* Additionally, half of the island-type distractors contained inanimate *wh*-phrases, providing a contrast with the animate *wh*-phrases in the experimental items. Across the entire item set, an equal number of questions began with *who, what,* and *which*-phrases. This was in order to remove any experiment-internal bias for processing one type of interrogative faster than another. After reading the question, subjects selected an answer from a set of alternatives provided. Of the three possible answers presented to them, one was correct (*the employee with poor sales*), another was lexically and syntactically similar (*the employee with poor hygiene*), and the third option differed drastically (*the cashier who stole money*).

Twenty subjects participated in this second study. All participants were Stanford University undergraduates who received course credit for their participation.

### 6.2 Results

The regions of interest for this experiment lie within the embedded *wh*-clause, as the left edge of this constituent constitutes the point when a proposed island boundary has been crossed. We focus in particular on reading times around the embedded verb where retrieval of the filler-phrase takes place, e.g. *dismissed* in (47).

The results verify that *which*-N phrases lead to faster reading times inside the embedded *wh*-clause. At the embedded verb, there is a (nonsignificant) trend for faster reading times in the WHICH condition ($t_1(19) = 1.619, p = .122, t_2(23) = 2.048, p = .052$). When the spillover regions for the verb are considered, however, the effect of *wh*-phrase complexity is observable and statistically significant. Considering the verb and the first word of the subsequent modifying material, the difference between *which*-N phrases and *who* is significant by both subjects and items ($t_1(19) = 2.127, p < .05; t_2(23) = 2.615, p = .015$). This difference persists for three words after the verb, as pictured in Figure 5 – reading times in the BARE condition remain slower for three words after the subcategorizing verb.

Taking the verb together with the next three words shows a highly significant effect of complexity ($t_1(19) = 3.600, p = .002; t_2(23) = 3.669, p = .001$). This suggests that at least some (if not a large proportion) of the difficulty associated with retrieving the *wh*-object continues after the verb is processed. According to planned comparisons, the processing facilitation associated with the complex *which*-N phrases also eliminated any substantial difference between the island-violating WHICH condition and the BASELINE, which does not violate any putative constraints on extraction.

[FIGURE 5 ABOUT HERE]

At the subject of the embedded clause, there is a slight difference between the BARE condition and the WHICH condition. But this effect is not significant, either by subjects or by items, as shown in Table 2.

(TABLE 2 ABOUT HERE)
Question-answer response times reflect an advantage for the WHICH condition (F1(1,19) = 13.664, p = .002, F2(1,23) = 10.778, p = .003). Given a which-N phrase in these contexts, as opposed to a bare wh-item, subjects answer a question with fewer possible answers; however, the likelihood of lexical priming is quite strong, given the layout and presentation of the stimuli. The presence of the head noun in the WHICH conditions and in the answers permits a simple recognition operation to take effect, since the head noun appears in the answer as well. Hence, while the question-answer response times also suggest more efficient processing in the WHICH condition, this conclusion is unwarranted because of the experimental confound of lexical repetition in the question and answer. Question-answer accuracies did not vary significantly across conditions within the experimental items (BARE = 89.6%, SE = 2.21; WHICH = 88.5%, SE = 2.31; BASELINE = 90.6%, SE = 2.99).

6.3 Discussion

Filler-gap processing, according to the evidence, improves when the filler-phrase encodes more information. The processing advantage, however, does not begin immediately after the filler-phrase. The more complex wh-form initially causes slower reading, likely due to the added task of integrating the information from the head noun of the which-N phrase – essentially, the cost of building a more complex representation.

The data also show that processing improves so much that the overall difference between the non-island BASELINE and the WHICH condition disappears. This finding aligns with the intuitions expressed in the theoretical syntax literature that extraction out of island contexts improves with the specificity of the extracted element. The absence of a difference between the BASELINE and WHICH conditions is notable for another reason: some small but significant decrease in processing difficulty may be viewed as immaterial if processing difficulty remains at an extremely high level. Given the fact that reading times in the WHICH condition are essentially equivalent to those of the baseline, however, the evidence argues for a meaningful interpretation of the facilitation.

Because the BASELINE and WHICH conditions produce similar reading times, both significantly faster than the BARE condition, the perception of unacceptability for wh-island violations involving vague or nonspecific filler-phrases may reasonably have its origins in processing-related difficulties. The evidence at hand, therefore, suggests an alternative to the putative grammatical mechanisms used to explain the previously unsubstantiated contrasts. Specifically, to the extent that acceptability is derived from considerations of processing difficulty, the observed processing contrast stemming from the complexity of the extracted wh-phrase may be playing a large role in the perception of the acceptability of the entire sentence.

To test the hypothesis that acceptability improves where processing difficulty decreases, a separate acceptability study was run using the stimuli from the reading-time task, with one alteration. To remove the pragmatic oddity of decontextualized questions, the wh-islands were presented as embedded questions, as in the following modified version of (47):

(48) a. Only a few individuals repeated who Albert learned whether we dismissed after the annual performance evaluations.
b. Only a few individuals repeated which employee Albert learned whether we dismissed after the annual performance evaluations.

All stimuli began with vague, quantified NP subjects, such as *some people* or *no one*, where no particular type of individual is named, under the assumption that such NPs would incur fewer processing costs than naming a specific individual or type of individual. Participants were instructed to rate how natural the examples sounded as sentences of English on a scale of 1-7 (7 being perfectly natural). As depicted in Figure 6, the more complex *wh*-phrases significantly improve judgments of acceptability ($F_1(1,15) = 15.964, p = .001$; $F_2(1,19) = 14.428, p = .001$).

This study therefore verifies that acceptability judgments fall where processing difficulty significantly increases. The parallel between acceptability and processing difficulty in *wh*-islands echoes the results of the CNPC study just reviewed. Of course, this interpretation of the data could be turned on its head: the evidence could be interpreted as showing that grammaticality predicts processing difficulty. The issue of how to interpret the relationship between acceptability and processing results is fundamental, but before we address it we must consider a final island context.

### 7 Experiment III: Adjunct Extraction

The third experiment we discuss here addresses the referentiality of the displaced element. This island study supplements the previous investigations by examining whether effects based on complexity are restricted to referential arguments. Cinque, Rizzi, and others have interpreted differential acceptability in island contexts as a function of referentiality. According to such theories, nonreferential adjuncts that differ in syntactic and semantic complexity should not produce the same effects that have been observed to distinguish putatively referential (*which*-N phrases) and nonreferential (*who*) phrases.

In contrast, a processing-based explanation for these acceptability differences avoids placing boundaries on the types of syntactic entities for which the relevant cognitive constraints are relevant. To the extent that these processing differences derive from general resource considerations, e.g. working memory limitations, referentiality should not be a necessary precondition for the observation of these complexity-based contrasts. Hence, we predict that a similar sort of processing facilitation should occur for more complex filler-phrases, even when these phrases constitute nonreferential adjuncts. To test this hypothesis, we consider here dependencies involving adjunct phrases extracted out of *wh*-islands.

Note that this is not equivalent to claiming that arguments and adjuncts (or referential vs. nonreferential arguments) should behave identically with respect to any one cognitive constraint.
7.1 Materials & Participants

Twenty-four adjunct extractions from wh-islands were employed in this study. Stimuli were systematically varied in terms of the amount of information contained in the extracted temporal adjunct phrase. In the BARE condition, the temporal adjuncts contained only two words, either how long or how often. The LONG condition had longer temporal adjuncts of at least three words and as many as eight, as exemplified in (49) below. As in the previous experiment, a baseline condition containing a different lexical complementizer, that, was also included in the study as a means of evaluating differences between the two main conditions of interest. The baseline condition always contained the shorter adjunct phrase that appeared in the BARE condition.

In contrast with the previous two experiments, subjects were instructed not to answer the comprehension question, but to indicate whether the question could be answered, given the information stated in the context sentence. Half of the total items appearing in this experiment, including half of the twenty-four adjunct extractions, presented questions which could not be reasonably answered given the preceding text. For instance, in the item shown in (50), none of the three versions of the question can be answered, since the preceding text makes no mention of how long the children played during the afternoon recess. Thus, after each question, the subjects saw the prompt, Is it possible to answer the question?, and were instructed to provide a negative response when the preceding text did not contain the necessary information to answer the question. Subjects were informed during the training session of this experiment that each item did have a correct answer, and consequently received negative feedback if they answered the question incorrectly. Prior to the presentation of the actual experimental stimuli, subjects became familiar with this task via a practice session with eight items.

(49) Julie discerned that the survivor had managed to stay alive for eight days after the crash in the harsh conditions.

BARE: How long did Julie observe whether the passenger had survived in the unbelievably harsh conditions?
LONG: For what period of time after the crash did Julie observe whether the passenger had survived in the unbelievably harsh conditions?
BASELINE: How long did Julie observe that the passenger had survived in the unbelievably harsh conditions?

(50) Andrew overheard the daycare staff discussing how they wanted to get away from the children for a few hours.

BARE: How long did Andrew hear whether the children had played during the daycare’s afternoon recess?
LONG: How many hours did Andrew hear whether the children had played during the daycare’s afternoon recess?
BASELINE: How long did Andrew hear that the children had played during the daycare’s afternoon recess?

This methodology encourages participants to read the comprehension questions carefully.
Generally speaking, subjects read comprehension questions faster than the preceding text, partly due to a relatively high degree of lexical overlap and the predictability of upcoming constituents. The methodology employed in this experiment consequently dissuades participants from relying on predictability and repetition in reading and answering the comprehension question itself.

Twenty-eight Stanford University students were paid $10 for their participation in this study.

7.2 Results

As in the previous experiment, we concentrate here on the reading times within the embedded clause. The data identifies two particular word regions where the experimental conditions create significantly different reading-times: (1) at the complementizer and (2) at the word after the complementizer. At the clause boundary, the complementizer \textit{whether}, appearing in both the \textsc{Long} and \textsc{Bare} conditions, results in faster reading times than the complementizer \textit{that} in the \textsc{Baseline} condition (\textsc{Bare-Baseline}: $t_1(27) = -2.930, p < .01; t_2(23) = -2.522, p = .019$; \textsc{Long-Baseline}: $t_1(27) = -6.325, p < .001; t_2(23) = 3.200, p < .01$). One potential explanation for this strong effect concerns the syntactic ambiguity of \textit{that}, which can initially be parsed as a determiner, an indexical NP, or a complementizer. This syntactic ambiguity therefore may lead to a temporary slow-down in parsing. The relative difficulty of the \textsc{Baseline} condition at the complementizer immediately disappears, however, at the next word and this condition ultimately leads to the fastest reading times throughout the embedded clause, as shown in Figure 7.

![FIGURE 8 ABOUT HERE](image)

The other main effect is found at the word after the complementizer, which was always the determiner \textit{the}. Here, the \textsc{Bare} condition produces significantly slower reading times than the \textsc{Long} condition by nearly 50 milliseconds ($t_1(27) = 3.484, p = .002; t_2(23) = 3.513, p = .002$). Here the baseline patterns with the \textsc{Long} condition – reading times for the baseline at the first word of the embedded clause are substantially faster on average than those for the \textsc{Bare} condition. Notice that the reading time difference cannot be simply ascribed to the number of words read previously, i.e. string position, as the \textsc{Baseline} and \textsc{Bare} conditions contain the same number of words.

![TABLE 3 ABOUT HERE](image)

In contrast with the previous two experiments, no effect of complexity is observed at the embedded verb. In fact, after the first word of the embedded clause, the \textsc{Long} condition is not processed significantly faster than the \textsc{Bare} condition at any one particular word inside the embedded clause. However, considering the entire embedded clause (except for the final word of the clause, where end-of-sentence wrap-up effects occur), the \textsc{Long} condition produces significantly faster reading times than the \textsc{Bare} condition ($t_1(27) = 2.356, p = .026; t_2(23) = 1.648, p = .113$). Of course, much of this effect stems from the first word itself. Additionally,
while reading times for the entire embedded clause are faster for the baseline, as compared to the BARE condition (t₁(27) = 3.218, p < .01; t₂(23) = 2.518, p = .019), the baseline did not produce significantly faster reading times in the embedded clause than the LONG condition (ts < 1). Reaction times did not differ significantly by condition (BARE = 867.45, SE = 38.19; BASELINE = 843.83, SE = 43.46; LONG = 819.45, SE = 33.57). As in the previous studies, there were no significant comprehension accuracy differences across conditions (LONG: 86.16%; SE = 2.31; BARE = 84.82%; SE = 2.40; BASELINE: 84.82%, SE = 2.40).

7.3 Discussion

As in the experiments discussed previously, greater syntactic and semantic complexity in dislocated adjunct phrases significantly facilitates subsequent processing. At the embedded clause boundary, the reading time measures indicate a highly significant advantage for the complex adjuncts. The effect of complexity, according to this evidence, operates independently of referentiality and also appears to be generally insensitive to the argument-adjunct distinction.

These facts consequently argue against an explanation that is based solely on the notion of referentiality. If temporal adjunct phrases refer to nonreferential entities, then varying the complexity of said phrases should not affect the ability to extract them out of islands, according to syntactic theories of relativized movement. The results indicate, however, that adjunct phrases are subject to some of the same principles of sentence processing as argument phrases.

While showing that adjunct dependencies are similar to argument dependencies, this study does not address why displacement of adjuncts out of wh-islands is less acceptable than displacement of arguments. A complete answer to this question goes beyond the scope of this paper; however, there are a number of reasons to suspect that this difference is attributable to processing-based differences as well: (1) retrieval cues carried by verbs with missing arguments (e.g. subcategorization and thematic role information) may be better suited for recovering arguments than adjuncts; (2) displaced arguments can provide indirect evidence for the retrieval site in the form of a missing obligatory constituent, but displaced adjuncts do not provide such evidence since they are optional; (3) fronted adjuncts will often be able to modify intervening verb phrases along the filler-gap path, thus raising the probability of a subsequent reanalysis. Beyond these issues, the mental representations associated with adjuncts may be generally harder to retrieve. Hence, it is reasonable to assume that adjunct dependencies present an assortment of processing difficulties that are absent in argument dependencies.

There are clearly a number of open areas of investigation for the processing of adjunct dependencies, but this third experiment clearly confirms that adjunct dependencies are also sensitive to the complexity of the dislocated phrase. As in the first two experiments, more complex filler-phrases produce faster processing inside the syntactic island. These results are not easily explicable under accounts that state movement constraints on the basis of referentiality or other categorical divisions among phrase types.

12Note that these relatively fast response times are due to the fact the actual question prompt was the same for all items, which essentially obviates the need to read the question.
8 Performance vs. Competence

These empirical investigations into the processing of island constraint violations converge on the conclusion that where processing difficulty increases, acceptability decreases. Crucially, the results suggest that these differences in processing are not insubstantial fluctuations – they can effectively eliminate the difference between islands and non-islands. One interpretation of these facts – the one that we adopt here – is that processing differences are influencing perceptions of acceptability. A competing alternative, however, is that it is the grammar itself (competence-based factors) which decides the degree of processing difficulty in these FDGs.\footnote{Yet another possibility is that the parallel results are merely coincidental and do not reflect a relationship between the two response types. Given that these parallels have been replicated in numerous contexts, including islands and non-island environments, as well as the well-motivated belief that acceptability judgments are influenced by processing, we consider it unlikely that these parallels are accidental.}

On such an account, less acceptable or grammatical constructions would lead to greater processing difficulty. In this section, we consider a number of points that favor the conclusion that processing effort underlies the acceptability differences, rather than the other way around. Taken together, these arguments strongly favor a causal story that begins with processing, while a number of theoretical difficulties arise if it starts with grammar.

First among these points is this: a processing-based explanation appeals to the existence of independently motivated factors, some of which may not even be specific to language. Locality or distance-based effects, referential processing load, and other relevant processing factors have been identified and substantiated outside the domain of island constructions as reliable predictors of processing load. They are, in other words, a necessary part of any general theory of sentence processing. In this sense, the explanation is ‘cost-free’ – no new, arbitrary, and language-specific constraints have to be introduced.

Additionally, the complexity-based effects observed here are found in uncontroversially non-island contexts. For instance, even in fully grammatical interrogatives like those in (51), reading time measures are faster at the retrieval site (e.g. record below) when the sentence begins with the more complex which-\(\overline{N}\) phrase (Hofmeister, 2007):

(51) a. Which album did the musician that Robert saw record with two popular blues guitarists?
   b. What did the musician that Robert saw record with two popular blues guitarists?

The existence of these complexity effects in non-island contexts poses a problem for any grammar-based account. If grammaticality differences are to explain the processing differences in islands, then some secondary explanation must be invoked to account for the same processing differences in syntactic contexts which do not contain violations of any known grammatical constraint. Perhaps nothing beyond considerations of elegance and good taste stands in the way of a grammarian who chooses to postulate intricate, noncategorical, and relativized rules that state that dependencies with more complex fillers are more acceptable than dependencies with less complex fillers. For that matter, a grammar of long-distance dependencies might also stipulate that shorter dependencies are more acceptable than longer dependencies.
In contrast to an explanation couched in terms of processing, a grammar-based characterization of the amply documented, pervasive graded acceptability of sentences with FGDs would therefore require some highly specialized linguistic machinery. This may include FGD-specific constraints like Subjacency, as well as a system for calculating fine-grained acceptability differences. Notably, such a calculator would in all likelihood serve no function other than to tally the result of aggregating grammaticality violations. Additionally, such a system must also provide for the possibility that the interaction of grammatical constraints can lead to interactions in sentence processing.

Judgments surrounding islands are also known to vary widely across and even within individuals (Braze, 2002). Some individuals seem fairly accepting of some kinds of islands, while others reject many more tokens of island violations. This type of variation in acceptability judgments emerges naturally on the processing account of islands, since individuals are known to differ significantly from one another in terms of working memory capacity (Just & Carpenter, 1992). The same individual, in fact, may have more or fewer resources available, depending upon factors such as fatigue, distractions, or other concurrent tasks. This can theoretically account for why the same individual can perceive islands differently over time. Unlike a performance analysis, a competence account appears fundamentally incapable of modeling such differences.

Finally, if the goal of linguistic inquiry is to explain human behavior, the performance-based view accounts for islands as the byproduct of general principles of cognition, while the competence-based view merely describes the behavior in ad hoc terms. That is, the competence-based theory, by its very nature, offers no insight into why islands exist. But a perspective on islands that depends on the findings of psycholinguistic research links the acceptability of islands to factors that generally affect sentence processing and acceptability. In other words, a performance approach not only captures the variation, but it explains why it exists in the first place.

While these arguments favor the conclusion that processing factors are responsible for much of the acceptability variation tied to islands, they do not rule out the possibility that the nature of grammar itself is responsible for some of the variation. The behavioral data and acceptability judgments reported here do not preclude the role of grammar in shaping the data. Indeed, it is impossible to prove, given the current state of our knowledge of both grammar and processing, that grammar has no part in creating the observed processing differences. But the simplification of the grammar that the processing perspective allows (e.g. the elimination of the Subjacency Condition), taken together with its reliance on independently motivated properties of language processing, make it by far the more attractive hypothesis.

9 Conclusion

The grammatical constraints that have been proposed to account for syntactic islands are almost uniformly complex, arbitrary, and language-specific. They constitute attempts to express intricate and highly specific limitations on just a subset of the linguistic dependencies possible in natural language. They are arbitrary in the sense that they bear no relationship to other constraints, emanate from no general principles of language, and have no relevance or parallel
outside language. In short, syntactic island constraints mark islands as special, even within the domain of language, and even more particularly, within the domain of linguistic dependencies. Consequently, island constraints offer little insight into anything about language or cognition, except islands themselves.

At the same time, the island constraints that have been proposed by linguists (at least those we have examined here: the Subjacency Condition, the Wh-Island Constraint, and/or the Complex Noun Phrase Constraint) face serious empirical difficulties. Stated as categorical constraints on dependencies, syntactic formulations of island constraints have been counterexemplified on countless occasions over the past half century. Within and across languages, nearly every proposed structural island constraint has been shown to be violable in at least some circumstances. Judgments have also been shown to vary on the basis of factors that do not alter the structure upon which the island constraint is based. Historically, much of the variation surrounding judgments of critical island-related data has been set aside as ‘exceptional’ or irrelevant. Without such uncritical and unmotivated data triage, which we have surveyed in detail, most syntactic theories of islands would have been falsified long ago.

Our approach here attempts to legitimate this variation and to find a more adequate means for analyzing it. In so doing, we have proposed a way of eliminating theoretically central island constraints from grammar. This welcome simplification of the theory of grammar interacts with independently motivated cognitive constraints to predict observed contrasts in reading time and acceptability judgments. This approach, which acknowledges the role of cognitive constraints on perceptions of acceptability, has been generally underrepresented in the syntactic literature. Despite the early recognition that performance-related factors can influence acceptability, very few syntactic analyses (since Chomsky & Miller 1963) have argued for theoretically relevant, performance-based distinctions. In essence, a major explanatory tool for linguistic behavior has been left unused. Given the prospect of accounting for a substantial amount of island variation using this tool, it is quite possible that a similar strategy can be applied to other cases of judgment variation. Particularly when the data surrounding a grammatical construction exhibit a large amount of variation and when nonstructural choices have major effects on acceptability and other behavioral measures, appeals to constraints on sentence processing may ultimately offer the most elucidating and economical explanation.

10 Appendix

10.1 Experiment I Stimuli

1. He knew which country who Emily heard (a/the rumor(s)) that we had invaded due to increased political instability.
2. She discovered which passage what Jacob read (a/the allegation(s)) that they had copied into the final written report.
3. I saw which convict who Emma doubted (a/the report(s)) that we had captured in the nationwide FBI manhunt.
4. She remembered which article what Michael denied (a/the suggestion(s)) that they had plagiarized in order to sound intelligent.
5. He forgot which song what Jessica reiterated (a/the contention(s)) that we had stolen from the original German composer.

6. I verified which patient who Chris held (a/the conviction(s)) that they had cured with the new experimental treatment.

7. He realized which prisoner who Ashley countered (a/the belief(s)) that we had interrogated without regard to international law.

8. She wondered which company who Matthew confirmed (a/the suspicion(s)) that they had sued for its unethical accounting practices.

9. I researched which student who Amanda conveyed (a/the threat(s)) that we would reject despite an outstanding academic record.

10. She guessed which client who Joshua disputed (a/the notion(s)) that they had defended before the federal appeals court.

11. He insinuated which actor who Jennifer overheard (a/the comments(s)) that we had arrested for drunk driving last night.

12. I acknowledged which novel what David expressed (a/the worry(ies)) that they would ban due to its racy content.

13. He indicated which concert what Sarah answered (a/the objection(s)) that we had canceled unnecessarily because of a disagreement.

14. She understood which intern who Daniel envisioned (a/the prospect(s)) that they would hire for the emergency room position.

15. I surmised which agency who Erin echoed (a/the complaint(s)) that we had overcharged for a routine financial report.

16. She testified which supervisor who James established (a/the expectation(s)) that they would fire because of his lewd behavior.

17. He determined which project what Nicole shared (a/the intuition(s)) that we would complete with extra time to spare.

18. I learned which route what Andrew proposed (a/the hypothesis(es)) that they had used to cross the mountain range.

19. He published which structure what Brittany announced (a/the plan(s)) that we would build to replace the condemned building.

20. She perceived which river what Robert addressed (a/the fear(s)) that they had polluted with dangerous levels of toxins.

21. I recorded which dictator who Heather contested (an/the assertion(s)) that we had supported in his rise to power.

22. She uncovered which quality what John debated (a/the perception(s)) that they had lacked in the home loan application.

23. He confirmed which student who Elizabeth believed (a/the charge(s)) that we had suspended due to their poor grades.

24. I specified which car what Ryan considered (a/the demand(s)) that they should recall because of failed safety tests.

25. He divulged which base what Megan recommended (a/the proposal(s)) that we should abandon in order to minimize casualties.

26. She recalled which leak what Joseph appreciated (a/the comment(s)) that they would repair within a week from now.
27. I proved which mineral what Melissa relayed (a/the message(s)) that we had identified on the surface of Mars.
28. She admitted which community who Brandon printed (a/the warning(s)) that they would forget in the disaster relief effort.
29. He confided which species what Amber verified (a/the theor(y)/(ies)) that we would discover with enough time and energy.
30. I asked which project what Justin recognized (a/the concern(s)) that they should coordinate after a series of disasters.
31. He investigated which election what Lauren repeated (a/the claim(s)) that we had rigged in favor of the Democrats.
32. She decided which company who William concealed (a/the sign(s)) that they had ruined with numerous illegal takeover attempts.
33. I resolved which spacecraft what Rachel noted (a/the signal(s)) that we had lost due to an insulation problem.
34. She perceived which tax what John protested (a/the request(s)) that they should pay for the next ten years.
35. He studied which resource what Danielle confessed (a/the feeling(s)) that we had depleted over many years of mismanagement.
36. I noticed which player what Nick disregarded (a/the comment(s)) that they would lose because of a serious injury.

10.2 Experiment II Stimuli

1. Kathy wondered if her friends consulted the doctor from New Madrid at the hospital last night.

Who\Which doctor did Kathy wonder whether they consulted at the hospital last night?

2. Nathan mentioned that the generals promoted the sergeant from western Tennessee during the war with Iraq.

Who\Which sergeant did Nathan mention whether they promoted during the war with Iraq?

3. Winston stated that the prosecutors trained the assistant who needed practice shortly before the trial began.

Who\Which assistant did Winston state whether they trained shortly before the trial began?

4. Sarah wondered if the voters elected the senator from Portland, Oregon in spite of the scandal.

Who\Which senator did Sarah wonder whether they elected in spite of the scandal?
5. Stephen confirmed that the radicals released the prisoner from New Zealand just after the demands were met.

Who\Which prisoner did Stephen confirm whether they released after the demands were met?

6. Brandy pondered if the investigators identified the suspect in Wednesday’s robbery at the station near Scranton.

Who\Which suspect did Brandy ponder whether they identified at the station near Scranton?

7. Jason learned that the doctors examined the patient with back pain very quickly at the clinic.

Who\Which patient did Jason learn whether they examined very quickly at the clinic?

8. Charlotte pondered if the delegates nominated the candidate who was liberal at the recent Republican convention.

Who\Which candidate did Charlotte ponder whether they nominated at the recent Republican convention?

9. Marvin mentioned that the violinists accompanied the pianist who played Beethoven at the concert on Sunday.

Who\Which pianist did Marvin mention whether they accompanied at the concert on Sunday?

10. Maureen speculated that the TAs tutored the student who was failing before the final physics exam.

Who\Which student did Maureen speculate whether they tutored before the final physics exam?

11. Albert learned that the managers dismissed the employee with poor sales after the annual performance evaluations.

Who\Which employee did Albert learn whether they dismissed after the annual performance evaluations?

12. Anna said that the members rejected the applicant from Yale University at the meeting about admissions.

Who\Which applicant did Anna say whether they rejected at the meeting about admissions?

13. Oscar stated that the detectives dispatched the officer with little experience after the tragic murder yesterday.
Who\Which officer did Oscar state whether they dispatched after the tragic murder yesterday?

14. Vera indicated that the Yankees retired the batter with two homers after the previous hitter doubled.

Who\Which batter did Vera indicate whether they retired after the previous hitter doubled?

15. Victor announced that the defendants intimidated the witness who was twenty with numerous threats of violence.

Who\Which witness did Victor announce whether they intimidated with numerous threats of violence?

16. Matthew verified that the landlords evicted the tenant who wasn’t quiet to satisfy the other residents.

Who\Which tenant did Matthew verify whether they evicted to satisfy the other residents?

17. Susan disclosed that the parents adopted the child who was sick after making several failed attempts.

Who\Which child did Susan disclose whether they adopted after making several failed attempts?

18. Eric announced that the agents arrested the criminal who escaped yesterday at a motel in Ohio.

Who\Which criminal did Eric announce whether they arrested at a motel in Ohio?

19. Erin disclosed that the terrorists killed the hostage from the U.S. in a moment of panic.

Who\Which hostage did Erin disclose whether they killed in a moment of panic?

20. Thomas indicated that the professors taught the graduate who studied psychology for at least two years.

Who\Which graduate did Thomas indicate whether they taught for at least two years?

21. Thelma speculated that the jurors acquitted the defendant charged with arson after discussing all the evidence.

Who\Which defendant did Thelma speculate whether they acquitted after discussing all the evidence?
22. Ryan verified that the zealots followed the leader of the cult to the islands without forethought.

Who\Which leader did Ryan verify whether they followed to the islands without forethought?

23. Rachel said that the cornerbacks sacked the quarterback for the Seahawks multiple times during the game.

Who\Which quarterback did Rachel say whether they sacked multiple times during the game?

24. Crystal confirmed that the teams rescued the survivor with severe injuries from the wreckage without difficulty.

Who\Which survivor did Crystal confirm whether they rescued from the wreckage without difficulty?

10.3 Experiment III Stimuli

1. Katherine was informed that the salesman often only got four hours of rest per night because of his insomnia.

(How long\How many hours per night) did Katherine find out whether the salesman had slept on account of his schedule?

2. Hilary appreciated the fact that the foreman was going to keep talking about his personal life until the work day was over.

(How long\For how much of the day) did Hilary understand whether the foreman would work before finally leaving for home?

3. Jack was aware that, twice a day, the sheriff had stopped by to gather additional information about the burglary.

(How often\How many times a day) did Jack know whether the sheriff had come to inquire about the burglary?

4. Julie discerned that the survivor had managed to stay alive for eight days after the crash in the harsh conditions.

(How long\For what period of time after the crash) did Julie observe whether the passenger had survived in the unbelievably harsh conditions?
5. Stephen was informed that the representative had gone missing for almost three hours prior to the committee meeting.

(How long\How many hours) did Stephen learn whether the representative had talked at the executive committee meeting?

6. Meghan received information that the patient was working out five days a week without losing any weight.

(How long\How many days per week) did Meghan discover whether the patient had exercised without seeing any weight loss?

7. In her graduate research, Lily looked into the unusual phenomenon of a group of Democrats voting Republican for 3 or more consecutive elections.

(How long\For how many consecutive elections) did Lily ascertain whether the Democrats had voted Republican while writing her dissertation?

8. Mark claimed that his students had spent half of the quarter editing haikus instead of discussing modern poetry.

(How long\What part of the academic quarter in total) did Mark state whether his students had been reading through the novels he assigned?

9. Mark claimed that his students had spent half of the quarter editing haikus instead of discussing modern poetry.

(How long\For how many months) did Hilary doubt whether some people would fish during the official fishing season?

10. Erica’s article determined that her school’s athletes had consumed the right amount of protein only once a week.

(How often\How many days per week) did Erica establish whether the athletes had eaten the recommended amount of protein?

11. Ethan concluded that one official had information about the modified bus routes for at least a few weeks.

(How long\For what length of time) did Ethan assess whether the official had known about the recent controversial firings?

12. Zack informed the others that the guard had not been coming to work punctually on Sunday mornings during the last 6 months.
(How long\For how many months) did Zack report whether the guard had noticed some sus-
picious activity taking place?

13. Kevin allowed one of his employees to spend two and a half weeks in the Bahamas
for her outstanding performance.

(How long\For how many weeks) did Kevin decide whether the employee could go to va-
cation in the Bahamas?

14. Morgan mentioned that the customer had lingered for five hours over the decision to
buy the computer.

(How long\For how many hours) did Morgan indicate whether the customer had got into
an argument with her?

15. Dana left her car with the mechanic for a week in July, before she left for a vacation
on Cape Cod.

(How long\For how many days in the summer) did Dana speculate whether the mechanic
had worked on fixing the broken engine?

16. Leslie noted that it was only for the last three hours before daylight that the lioness
prowled without making a sound.

(How long\For how many hours before daylight) did Leslie perceive whether the lioness had
moved without making a single sound?

17. Renee did not believe that the reporter would be interested throughout the entire show,
since it went on for so long.

(How long\For how much of the event) did Renee confirm whether the reporter would stay to
speak with the celebrities?

18. Alex confessed that, for at least the first week of the school year, the tenant was happy
with the living arrangement.

(How long\For how many weeks in the school year) did Alex say whether the tenant had
felt happy with the living arrangement?

19. Andrew overheard the daycare staff discussing how they wanted to get away from the
children for a few hours.

(How long\How many hours) did Andrew hear whether the children had played during the
daycare’s afternoon recess?

20. Bill couldn’t help but see that his officemate was in a meeting with the boss for several hours after the news of the merger.

(How long \ How many hours) did Bill notice whether his colleague had argued about the wisdom of merging?

21. Peter found in his old notes that the garden had blossomed for two weeks less than their normal three months because of a drought last year.

(How long \ How many weeks less than normal) did Peter note whether the flowers had bloomed during last spring’s dry spell?

22. Jennifer’s father told her that in 1990 he took five years off to learn some new skills before starting a different career.

(How long \ How many years) did Jennifer verify whether her father had traveled before returning to new work?

23. One day, Mark was informed that his favorite painting at his great uncle’s house had been there for a hundred years.

(How long \ How many years) did Mark determine whether the painting had hung at his great uncle’s house?

24. Jane knew that her nephew had been crashing at her parents for a couple of weeks, judging from the mess in the living room.

(How long \ How many weeks) did Jane realize whether her nephew had stayed at her parents’ now-messy house?

References


Allwood, J. (1976). The complex NP constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. In J. Stillings (Ed.), University of Massachusetts Occasional Papers in Linguistics, Volume II. Amherst, MA: University of Massachusetts at Amherst.


Rizzi, L. (2000). Reconstruction, Weak Island Sensitivity, and Agreement. (Ms. Università di Siena)


Figure 1: Residual reading times at complementizer in experiment I

Figure 2: Residual reading times at auxiliary + verb in experiment I
Figure 3: Mean judgment ratios of complex noun phrase constructions

Figure 4: Mean judgment ratios of complex noun phrase constructions
Figure 5: Residual reading times in experiment II, ranging from first word after \textit{w}h-phrase to three words after retrieval site. Error bars show one standard error.

Figure 6: Mean judgment ratios of embedded \textit{w}h-island violations: BARE = bare \textit{w}h-phrase; WHICH = \textit{w}hi\textit{ch}-\text{\textbar}N phrase
Figure 7: Residual reading times at first word after complementizer in experiment III
Table 1: Effect of complexity by region inside embedded clause in experiment I.
<table>
<thead>
<tr>
<th>Word</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual RT</td>
<td>BARE -47.32 (3.81)</td>
<td>WHICH -67.52 (5.27)</td>
<td>BARE -34.19 (4.39)</td>
<td>WHICH -33.50 (5.28)</td>
<td>BARE -48.56 (4.55)</td>
</tr>
<tr>
<td></td>
<td>WClH -56.88 (4.09)</td>
<td>WHICH -82.05 (5.22)</td>
<td>WClH -49.56 (4.66)</td>
<td>WHICH -59.48 (4.62)</td>
<td>WClH -69.83 (4.92)</td>
</tr>
<tr>
<td>Raw RT</td>
<td>BARE 251.19 (5.12)</td>
<td>WHICH 255.49 (5.69)</td>
<td>BARE 261.84 (5.12)</td>
<td>WHICH 265.46 (6.26)</td>
<td>BARE 259.76 (4.68)</td>
</tr>
<tr>
<td></td>
<td>WHICH 243.10 (4.72)</td>
<td>WHICH 242.63 (4.53)</td>
<td>WHICH 248.77 (5.05)</td>
<td>WHICH 242.22 (4.76)</td>
<td>WHICH 239.24 (4.44)</td>
</tr>
<tr>
<td>$t_1$</td>
<td>BARE 1.366</td>
<td>WHICH 1.619</td>
<td>BARE 2.369*</td>
<td>WHICH 3.479**</td>
<td>BARE 4.411***</td>
</tr>
<tr>
<td>$t_2$</td>
<td>WHICH 1.716 (.)</td>
<td>WHICH 2.048 (.)</td>
<td>WHICH 2.244*</td>
<td>WHICH 3.574**</td>
<td>WHICH 3.421**</td>
</tr>
</tbody>
</table>

Table 2: Effect of complexity (who vs. which-N) by region inside embedded clause in experiment II.
<table>
<thead>
<tr>
<th>Word</th>
<th>Region₁</th>
<th>Region₂</th>
<th>Region₃</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Residual RT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMPLE</td>
<td>-114.537 (10.61)</td>
<td>-12.371 (10.09)</td>
<td></td>
</tr>
<tr>
<td>COMPLEX</td>
<td>-137.786 (8.66)</td>
<td>-53.830 (6.11)</td>
<td></td>
</tr>
<tr>
<td>BASELINE</td>
<td>-80.126 (6.33)</td>
<td>-61.545 (5.33)</td>
<td></td>
</tr>
<tr>
<td><strong>Raw RT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMPLE</td>
<td>345.326 (11.09)</td>
<td>354.739 (11.29)</td>
<td></td>
</tr>
<tr>
<td>COMPLEX</td>
<td>324.032 (8.72)</td>
<td>311.442 (7.45)</td>
<td></td>
</tr>
<tr>
<td>BASELINE</td>
<td>310.113 (7.30)</td>
<td>301.442 (7.45)</td>
<td></td>
</tr>
<tr>
<td>(t_1)</td>
<td>SIMPLE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t_2)</td>
<td>COMPLEX</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Effect of complexity by region inside embedded clause in experiment III.