# Effects of processing on the acceptability of 'frozen' extraposed constituents

Philip Hofmeister, Peter Culicover, Susanne Winkler

*Abstract* In syntactic theory, 'freezing' refers to the idea that a constituent extraposed to a non-canonical position is resistant to extraction of any of its subconstituents (*What did Terry see a movie yesterday about ___ ?*). The unacceptability of such examples, compared to minimally different sentences without extraposition, has been claimed to be a result of a grammatical constraint on dependency formation, e.g. Ross' (1967) Frozen Structure Constraint. Here, we argue that the gradient pattern of acceptability associated with such examples is better explained in terms of processing complexity. Experimental evidence from a controlled acceptability task shows that the total penalty for extraction from extraposed constituents is predictable from the summed acceptability penalties independently linked to extraction and extraposition. As prior research links both of these displacement phenomena to increased processing costs, we conclude that a grammar of unbounded dependencies in English does not require a 'freezing' constraint for the explanation of these cases.[1]

## 1   Introduction

Dependencies in English, as in other languages, are theoretically unbounded in length. This means that there is no theoretical limit to the number of words that can separate a *wh*-word like *what* and the point at which it is thematically interpreted. The example in (1) could, in theory, continue on and on with more embedded complement clauses *ad nauseum* and not violate any rules of English grammar:

(1)     What did the police officer say that it was commonly known that it was illegal to claim that . . .

In light of this observation, there are a number of otherwise puzzling instances of ill-formedness that involve long-distance *wh*-dependencies. Some syntactic accounts, such as those of Chomsky (1964) and Ross (1967), propose that grammatical constraints account for the unacceptable status of items like (2-b), derived from the minimally different (2-a):

(2)     a.   Terry bought a book [s that explains string theory].
        b.  *What did Terry buy a book [that explains ___]?

Ross (1967), for instance, proposes a Complex NP Constraint that rules out extraction from a relative clause. The conclusion that there is such a grammatical constraint follows from the observation that questions in English are formed by moving the *wh*-phrase (in this case, *what*) to the front of the sentence. Since such a movement occurs in (2-b) and unacceptability results, a straightforward explanation is that a syntactic constraint renders extractions from such contexts ungrammatical. The research program that explores the nature of such constraints has been enormously productive, and has led to fundamental developments in syntactic theory (represented by such works as Chomsky 1973, 1977, 1981, 1986, 1995).

An alternative perspective builds on the observation that not all cases of unacceptability are due to ungrammaticality. This was articulated early on in the history of contemporary syntactic theory by Miller & Chomsky (1963). They observed that sentences with multiple center-embedding were structurally well-formed or grammatical, but difficult or impossible to process, and hence unacceptable. Subsequently, there have been a number of attempts to account for apparent ungrammaticality in terms of processing or learning difficulty, e.g. Bever (1970), Bever (2009), Clark & Haviland (1974), Deane (1991), Kluender & Kutas (1993b). For example, Jackendoff & Culicover (1972) offer a processing explanation for the intuition that $\overline{\text{A}}$ movements of indirect objects in English are mildly unacceptable, as in (3) & (4):

(3)  a.  Terry bought Robin an ice cream cone.
     b. ??Who did Terry buy __ an ice cream cone?

(4)  a.  Terry will send Robin a letter.
     b. ??Who will Terry send __ a letter?

Kluender (1992, 1998, 2004, 2005) and Kluender & Kutas (1993b) subsequently proposed that unacceptability attributed to certain apparent grammatical constraints on long-distance dependencies, viz. island constraints, were actually the consequence of processing difficulty. On the basis of reading time and acceptability data, a case has been made for processing accounts of a wider range of phenomena by Sag, Hofmeister and colleagues (Hofmeister 2007; Hofmeister & Sag 2010; Hofmeister, Jaeger, Sag, Arnon & Snider 2007; Hofmeister, Jaeger, Arnon, Sag & Snider 2013; Hofmeister, Staum Casasanto & Sag 2013; see also Chaves 2013; Christensen, Kizach & Nyvad 2013; Kim 2013 and, for an alternative perspective on these phenomena, see Phillips 2006; Phillips 2013; Sprouse, Wagers & Phillips 2012).

Our focus here is on explanations that appeal to 'freezing' as a grammatical constraint. In syntactic theory, freezing refers to the idea that a constituent displaced to a non-canonical position is resistant to extraction of any of its subconstituents (Wexler & Culicover 1980). A typical example of a frozen structure is a PP that has been extraposed. As (5-b) shows, extraction from such a PP causes a reduction in acceptability compared to (5-a) (the use of symbols like $t$ is merely to indicate the position where the *wh*-word is thematically interpreted, but it is not meant to endorse the syntactic or psychological validity of movement traces).

(5)  a.  He saw [a picture ___] yesterday [PP of Thomas Jefferson].
     b.  ?Who$_i$ did he see [a picture ___] yesterday [PP of t$_i$]?

The notion that some syntactic configurations produce freezing effects first appears in Ross (1967:305), and has been invoked subsequently to account for a substantial range of phenomena (see Corver 2006 for a review). Ross' (1967) version of the freezing constraint in (6) deals specifically with examples like those in (5):

(6)     The Frozen Structure Constraint (FSC): If a constituent C, where C is a clause or a prepositional phrase, has been extraposed from a noun phrase whose head is lexical, this noun phrase may not be moved, nor any element of C be moved out of C (pp. 160, 165).

From the perspective of linguistic theory, such a grammatical principle rules out extraction from extraposed structures. Nonetheless, such structures are also characterized by properties that complicate sentence processing, independently of the proposed structural constraint. In (5-b), for example, the processor is holding *who* in memory when it encounters *picture*. It has no information at that point that the gap is in a PP, and because of the extraposition, there is no information at *yesterday* that a modifying PP is upcoming that will modify the NP *a picture*. When *of* is encountered the processor must simultaneously link *who* to the gap, link this reconstructed PP to the extraposition site, and interpret the PP as a complement of *picture*. The question, therefore, is: Is the lack of acceptability in freezing configurations a matter of grammar per se, or could it have a processing explanation?

The present work builds on Culicover & Winkler (2012), who argue that the unacceptability of freezing constructions arises from an interaction between the grammatical structure and the mechanism by which the structure is interpreted in real time. Culicover and Winkler observe that not all such extractions are equally unacceptable; in fact, some are quite acceptable. For example, consider the following contextualized example from Huck & Na (1990) with contrastive emphasis on the sentence-final preposition:

(7)     I know Alger found letters in the file TO Chambers, certainly, but I'm not sure I can remember whom$_i$ he found letters in the files FROM $t_i$. [Huck and Na 1990:66 (39c)]

Building on such observations, Culicover and Winkler claim on the basis of intuitions that examples with the same syntactic structure with respect to freezing conditions vary in acceptability as a function of non-structural factors such as the distance between the extracted element and its corresponding gap.

Similarly, our hypothesis is that freezing effects emerge from the combination of independent sources of processing costs: extraction and extraposition. Filler-gap dependencies, a.k.a. extractions, introduce numerous costs, including the search for the correct point of retrieval and integration — the 'gap' — while processing intervening material (Wanner & Maratsos 1978; Deane 1991; King & Just 1991; Kluender 1991; Just & Carpenter 1992; Kluender & Kutas 1993a, 1993b; King & Kutas 1995; Gibson 1998; Hawkins 1999, 2004; Warren & Gibson 2002, 2005). In this sense, filler-gap dependency processing generally gives rise to a multi-tasking problem.

On top of this, freezing violations involve extraposition or the rightward dislocation of a constituent from its canonical position. Here, too, the available evidence suggests that extraposition is a source of processing difficulty. Like extraction, extraposition separates two syntactically and semantically related constituents. Levy, Fedorenko, Breen & Gibson (2012) demonstrate through self-paced reading experiments that extraposition of relative

clauses across VPs and PPs creates processing difficulty, compared to in-situ relative clauses. As discussed by Levy et al. (2012), there are various possible explanations for why extraposition creates comprehension difficulty (see also Müller 2004). However, our goal here is not to decide which of these best captures effects of extraposition, particularly since our only behavioral measures are acceptability judgments, which do not always capture differences in online sentence processing. Rather, we note that all of these views on extraposition converge on the prediction that processing sentences with extraposed constituents is harder than processing ones with in-situ constituents (but see the General Discussion for some speculations).

Freezing effects and the related grammatical proposals may strike some as being off the beaten path and narrowly concerned with peripheral data, particularly those dealing in non-transformational approaches to syntax. We believe, however, that the topic sheds light on an issue of quite general importance: how should judgments of unacceptability be analyzed and how can we adjudicate between such competing explanations of acceptability patterns (see, among many others, Marks 1968; Bever 1970; Pylyshyn 1973; Bever, Carroll & Hurtig 1976; Kluender 1991; Schütze 1996; Cowart 1997; Fanselow & Frisch 2006; Sprouse 2009; Hofmeister & Sag 2010; Featherston 2011; Sprouse, Wagers & Phillips 2012; Goodluck & Zweig 2013; Hofmeister et al. 2013; Phillips 2013)?

The approach we take here is to find independent evidence of general processing costs tied to a structure, see how much these costs can explain acceptability contrasts, and then ask what's left for a hypothetical grammatical constraint to explain. This method builds on other recent investigations into the source of acceptability contrasts based on considerations of additivity (Sprouse 2007; Sprouse et al. 2011; Sprouse, Wagers & Phillips 2012; Sprouse & Almeida to appear; Staum Casasanto, Hofmeister & Sag 2010). The basic logic runs as follows. Suppose two independent sentence processing costs each lower acceptability judgments, and the net acceptability drop when both processing costs are simultaneously present is no more than the summed independent acceptability effects. In this case, there is little reason to posit some additional mechanism to explain the unacceptability effects.

Using this Occam's razor logic, we arrive at the conclusion that there is little, if anything, for grammar to explain in the case of freezing effects. This conclusion derives from acceptability data showing that key features of freezing cause acceptability drops even when these examples do not violate any putative grammatical principle. Specifically, extraposition results in lower judgments, even in the absence of extraction, and extraction too lowers judgments independently of extraposition. We go on to show that the reduced acceptability of freezing cases is no more than expected from the summed, independent acceptability penalties for extraction and extraposition, leaving nothing to be explained by a grammatical constraint.

In the next section, we test how the key features of freezing — extraction and extraposition — independently affect judgments and how these effects summate. In the concluding section, we return to the issue of how to best interpret the difficulty associated with PP extraposition and the conclusions which can be drawn in cases of additive vs. super-additive effects on acceptability judgments.

4

# 2 Acceptability Judgments: Extraction and Extraposition

In this experiment, we cross manipulations of extraction and extraposition to assess whether the unacceptability of freezing violations is predictable from the acceptability decrement due to extraction plus that due to extraposition. The guiding assumption is that both extraction and extraposition give rise to processing costs which may systematically influence acceptability judgments. This hypothesis is motivated by a substantial body of evidence from self-paced reading and eye-tracking tasks showing that increasing the distance between syntactically-related linguistic units slows processing times at the point of integration (Ferreira & Henderson 1991; Gibson 1998; Gibson 2000; Grodner & Gibson 2005; Bartek et al. 2011). In contrast, grammatical accounts assume that freezing effects reflect something beyond the sum of the independent effects of extraction and extraposition.

For the purposes of our experiment, we assume that English grammar allows for the extraposition of the PP anywhere in the VP. That is, we take the traditional view that a PP can be dislocated from an NP and adjoined to the right edge of a VP (Ross 1967). Assuming that the PP can adjoin to the edge of any VP, the fact that extraposed PPs can appear before other constituents of VP is compatible with a recursive VP structure, e.g.:[2]

(8)　　a.　$[_{VP_1} [_{VP_2} [_{VP_3}$ V [NP PP] $XP_1$ ] $XP_2$ ] $XP_3$]
　　　　b.　$[_{VP_1} [_{VP_2} [_{VP_3}$ V [NP ___ ] $XP_1$ [PP] ] $XP_2$ ] $XP_3$]

Given a flat structure in the VP (e.g. Culicover & Jackendoff 2005), extraposition can be viewed as movement of the PP to the right of an XP sister of NP in (9-b), compared to the non-extraposed structure in (9-a).

(9)　　a.　$[_{VP}$ V [NP PP] $XP_1$ $XP_2$ $XP_3$]
　　　　b.　$[_{VP}$ V [NP ___ ] $XP_1$ [PP] $XP_2$ $XP_3$]

## 2.1 Materials

Our experimental materials varied with respect to whether or not there is extraction, as well as whether the critical PP is extraposed or not. Crossing these two factors leads to example items like the one below:

(10)　　a.　Kenneth revealed which President he overheard a nasty remark about earlier.
　　　　　　= [+ EXTRACTION, − EXTRAPOSITION]
　　　　b.　Kenneth revealed which President he overheard a nasty remark earlier about.
　　　　　　= [+ EXTRACTION, + EXTRAPOSITION]
　　　　c.　Kenneth revealed that he overheard a nasty remark about the President earlier.
　　　　　　= [− EXTRACTION, − EXTRAPOSITION]
　　　　d.　Kenneth revealed that he overheard a nasty remark earlier about the President.
　　　　　　= [− EXTRACTION, + EXTRAPOSITION]

---

[2]Note that it is movement per se that raises the possibility of a freezing explanation. The details would be different in an account that does not require movement of the PP at all. For example, Rochemont and Culicover (1997) take extraposition to be a relation between a predicate in VP and an NP antecedent. On such an account, the freezing explanation does not apply, but the phenomena covered in our experiment still require an explanation.
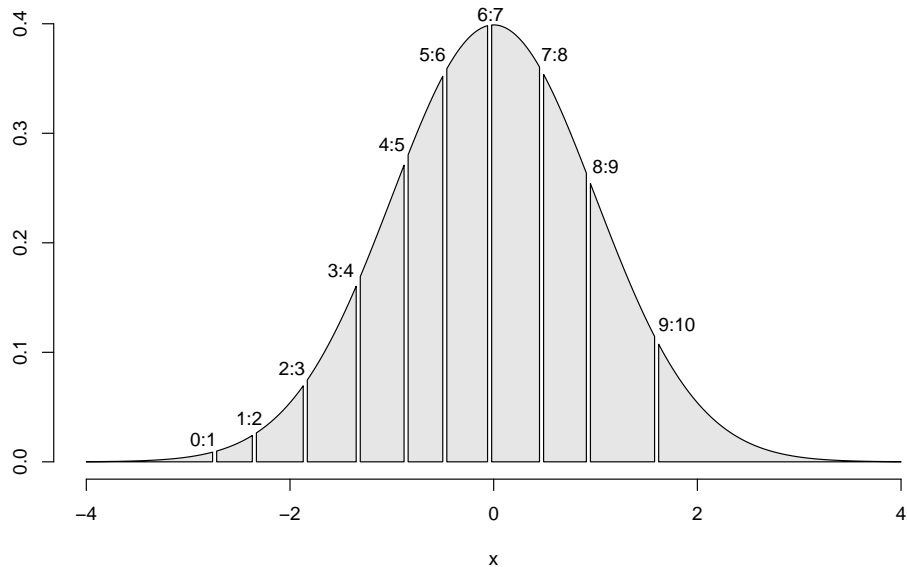
Figure 1: Acceptability thresholds for the critical materials in an ordered categorical probit model

This design allows us to determine how much extraposition and extraction independently lower judgments, and whether combining the two lowers judgments beyond what is expected on the basis of each independent factor. In our materials, the intervening phrases consisted of adverbials or prepositional phrases that modified the verbal predicate.

Participants read twenty-four such items, but only saw one version of each item. Prior to analysis, however, we removed the results from four items due to possible parses that did not involve extraposition (see Appendix for a list of the items utilized). That is, some item conditions which were designed to require a parse with extraposition possessed alternative readings without extraposition, e.g. *I was amused to find out which Russian author James read a book all night by*, which has an interpretation where the reading event took place next to a Russian author.

The experimental materials were presented along with forty-eight unrelated distracters. These distracters spanned the spectrum of acceptability from highly natural items (*Republicans and Democrats in the State Senate appeared close to a power-sharing deal*) to highly unnatural items (*Ransoms started in of out the tens of thousands dollars and have since climbed into the millions*). At least one filler appeared between every pair of critical trials and each survey began with three filler trials.

Finally, it is worth pointing out that we developed materials here that minimized the chances of readings without extraposition by interposing adverbials or other modifiers that could not combine with the extraposed PP. Without such deliberate formulation, extraposition of a PP often leads to alternative parses, as in the following constructed example:

(11)    Which topic did the candidate avoid answering a question in the allotted time about

6

|                    | Coefficient | Standard Error | t-value |
|-------------------:|:-----------:|:--------------:|:-------:|
| EXTRACTION         | −0.71       | 0.07           | −10.08  |
| EXTRAPOSE          | −0.41       | 0.07           | −6.22   |
| EXTRACT x EXTRAPOSE | 0.29       | 0.14           | 1.99    |

Table 1: LME summary for acceptability z-scores

> as best he could?

Here, the extraposed preposition could be combined with the preceding noun or the following adverbial. This is relevant not just for the creation of experimental materials but also relates to prior examples in the literature, such as the following from Ross (1967), where the extraposed PP can be parsed as the sister of *her teacher*:

(12)     Who is Ann going to send a picture to her teacher of, as soon as she gets home?

In other words, local attachment preferences will generally disfavor an extraposition reading. Such processing heuristics, therefore, will generally cause extraposition to be unexpected, creating even greater difficulty when the sentence forces an extraposition reading (which may cause extraposition to be even less expected, thus renewing the cycle; see Chaves 2013 for a similar account of subject islands). The point here is that sentences with extraposition often contain difficulties which we have attempted to minimize in our experiment, but our intent is not to dismiss these difficulties as irrelevant to the overall perception of sentences with extraction from extraposed constituents.

## 2.2   Participants & Procedure

We collected acceptability judgments via Amazon.com's Mechanical Turk marketplace. Gibson, Piantadosi & Fedorenko (2011), Sprouse (2011), and others have demonstrated that this means of collecting judgments produces results similar to traditional laboratory methods, while simultaneously providing the advantages of large participant numbers and relatively short data gathering times.

72 individuals completed the survey on Mechanical Turk. At the beginning of the experiment, participants identified their physical location and language background. To remove any incentive to lie about language background, participants were not excluded from the experiment on the basis of language history. The results from 4 participants who did not indicate English was their first language were eliminated prior to analysis. Participants received $1.50 for their participation.

Participants were instructed to judge the subsequent sentences in terms of naturalness on a 0-10 scale, 0 being extremely unnatural and 10 being extremely natural.[3] Following each sentence, we included a comprehension question to ensure that participants comprehended the materials. We planned to eliminate data from participants with comprehension accuracy rates below 75%; however, no participant scored below this threshold (mean accuracy = 95.6%).

---

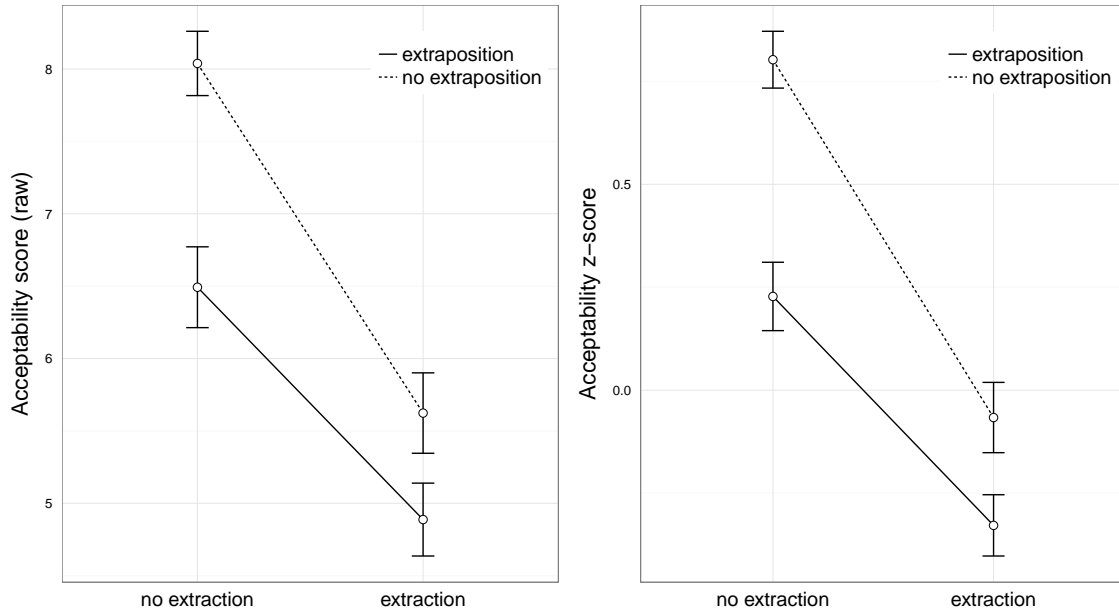[3]We opted for such a scale over the more standard 1-7 scale as a potential safeguard against floor effects.

Figure 2: Mean acceptability judgments by condition for raw judgments (left) and judgment z-scores (right); error bars show ± 95% CIs

Prior to analysis, we computed acceptability z-scores separately for each subject, based on the raw judgments from all experimental trials, including fillers. This takes into account different uses of the rating scale and also transforms the ordered categorical responses into linear data. As Figure 1 depicts, the cutpoints for the acceptability scale are roughly equally spaced, as determined by an ordered categorical response probit model, using the `clmm()` package via R (version 3.0.1), with fixed effect terms for extraction and extraposition and their interaction, as well as random intercepts for participants and items.[4] This feature of the data supports the idea that the data may be reasonably treated with standard linear regression techniques, particularly after transformation (see Gelman & Hill 2007, p. 123), i.e. the danger in transforming and treating ordered categorical response data as linear data arises most strikingly when the category levels are unequally spaced or only a limited range of possible values actually appears in the data, neither of which is problematic for this experiment.

We analyzed all acceptability z-score data by using linear mixed effects (LME) models with participants and items modeled as random variables (Pinheiro & Bates 2000; Baayen, Davidson & Bates 2008). Prior to analysis, all predictors were centered and higher order variables were based on these centered predictors. In addition, we utilized the maximum random effect structure possible. Thus, the random effect structure included random intercepts for participants and items, as well as by-participant and by-item random slopes for each factor and their interaction. Although models with nested random effect structures

---

[4]The probability distribution of the cutpoints reflects the fact that roughly half of the stimuli received judgment ratings of seven or above. Moreover, judgments of ten occurred 6.7 times more often than judgments of zero across all conditions.
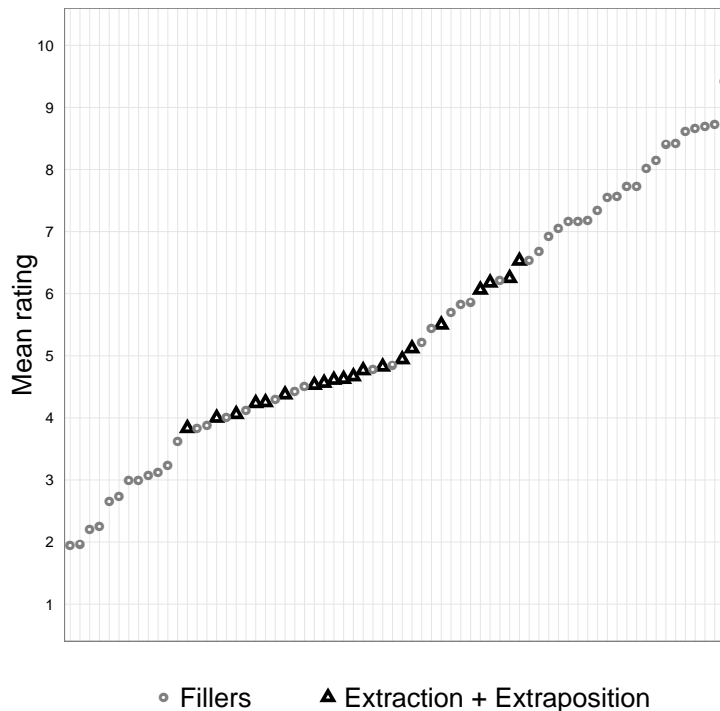
Figure 3: Mean raw acceptability judgments for fillers and conditions with extraction and extraposition ('freezing violations')

do not directly yield p-values, significance at the .05 level can be conservatively estimated for fixed effects coefficients with t-values which have absolute values at or above 2 (Baayen 2008; Baayen, Davidson & Bates 2008; Pinheiro & Bates 2000).

## 2.3 Results

As Figure 2 shows, extraposition lowers judgments significantly. Examples like (10-b) & (10-d) are judged to be worse than those like (10-a) & (10-c). Extraction also lowers judgments: sentences like (10-a) & (10-b) receive lower ratings than sentences like (10-c) & (10-d). However, extraposition is no worse in contexts with extraction than in contexts without. In fact, the opposite pattern is apparent in the results: extraposition has a weaker effect in the context of extraction, which accounts for the marginally significant interaction term (see Table 1).

In theory, a lack of super-additive penalties in any acceptability study may arise because judgments simply cannot go any lower, i.e. the bottom of the scale has been reached. However, in the current case, the evidence argues against such a floor effect. Not only was the grand mean of the condition with extraction and extraposition near the middle of the rating scale ($\mu = 4.89$, $SE = .13$), but 20 out of 48 fillers received lower mean ratings than this, and some quite substantially lower, e.g. the lowest mean rating was 1.94 ($SE = .21$). To illustrate a similar point, Figure 3 shows the mean rating of every filler in the experiment and the mean rating for all sentence versions with both extraction and extraposition. Even

the worst such 'freezing violation' receives a higher mean rating than a quarter of the overall fillers in the experiment.

## 2.4   Discussion

Extraction out of extraposed constituents does not lead to unacceptability lower than expected on the basis of the independent penalties attributed to extraction and extraposition. Each of these syntactic operations, which are fundamental to freezing effects, independently lowers judgments. Yet putting these two operations together, as the results show, does not result in any extra unacceptability. In fact, 'freezing violations' were effectively judged higher than anticipated on the basis of the independent effects due to extraction and extraposition, although this under-additivity is borderline.[5] Moreover, the pattern of our findings appears quite consistently across our items, as evidenced in Figure 4. Most items show a pattern of additivity or under-additivity, suggesting that our current results are not merely a consequence of a few atypical or unrepresentative items.[6] As there is no drop in acceptability beyond that predicted by the independent effects of extraction and extraposition, we are led to the conclusion that no other factors need to be invoked to account for freezing effects.

# 3   General Discussion

There are no direct tests for what makes a sentence unacceptable. Since many structural generalizations appear to have some functional purpose (e.g. case-, number-, and person-marking all serve to explicitly mark the relationship between constituents that can be arbitrarily far from each other), grammatical analyses often compete with functional or processing-oriented explanations of the same phenomena. Why a particular structure sounds unnatural could thus be due to grammatical principles, the processing costs the structure gives rise to, or both. This complicates any attempt to test whether processing costs rather than grammatical principles account for some acceptability contrast. In support of a processing-based analysis of freezing effects, our approach has been to suggest that fundamental features of freezing violations – extraposition and extraction – give rise to processing difficulty and that the costs associated with each of these plays a major role in how examples of freezing violations are judged. Our main finding, in this regard, is that the unacceptability of freezing violations is predictable from the independent penalties attributed to each of these key features: the combined penalty for extraction from extraposed PPs is no more than one would expect from combining their independent effects. If freezing is a special case of a more general situation, a grammatical principle like the FSC becomes

---

[5]Two items, #16 & #20, appear to be the main drivers of this under-additivity. In these cases, all conditions with extraction and/or extraposition were equally bad. Removing these items from the analysis roughly halves the interaction coefficient and associated t-value.

[6]As Figure 4 shows, while most items reflect an additive or under-additive pattern, a minority display a pattern consistent with super-additivity, e.g. #6, and extraposition by itself has larger effects for some items rather than others. This unsurprisingly implies that other features of the sentences in question have an important role to play, including features like the length of the interposing material, the identity and length of the extraposed preposition, how often the host NP appears with prepositions in general or a specific preposition, among other factors. In short, the acceptability of sentences with extraction from extraposed constituents undoubtedly depends upon more than the syntactic configuration.
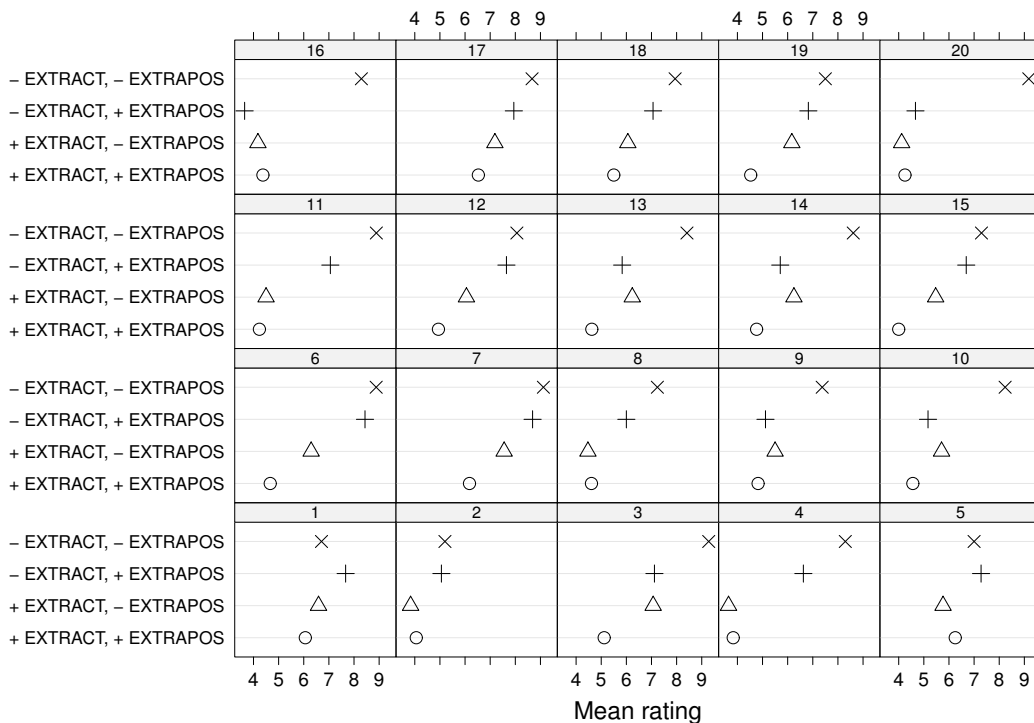
Figure 4: Mean raw acceptability judgments by item and condition

otiose and the grammar of long-distance dependencies can be simplified, i.e. extraposition plays no role in the constraints on long-distance dependencies.

At heart, we are brandishing Occam's razor: if two independent sentence-level features lower judgments, and the result of combining them is no more than the sum of their independent effects, there is no reason to postulate a further linguistic mechanism to account for the observed pattern (for similar reasoning, see Sprouse 2007; Sprouse, Wagers & Phillips 2012). We must emphasize, however, that additivity, under-additivity, and super-additivity are NOT, by themselves, diagnostic when it comes to acceptability judgments. At an abstract level, any particular pattern of judgments — under-additivity, strict additivity, and super-additivity — is technically reconcilable with the idea that grammar (or conversely, processing limitations) has some role to play in that pattern.[7] As reported by Hofmeister,

---

[7]It should be noted that under-additive effects are entirely possible when combining the costs of two temporally overlapping tasks. In the case of freezing violations, relating the preposition to the preceding NP and relating it to the extracted phrase occur nearly simultaneously, creating potential competition if particular stages of each task cannot be performed in parallel. Using the logic described by Pashler (1994) and Pashler & Johnston (1998), under-additive effects would be predicted if extraposition or extraction affect a stage that precedes the bottleneck. To see how this would work, suppose for the sake of illustration that discourse elements in memory can be retrieved in parallel, and that both extraposition and extraction affect retrieval latencies. Second, assume some later stage (integration, thematic role assignment, etc.) permits only one process at a time to proceed, i.e. it imposes a bottleneck. When the postponement caused by the bottleneck is longer than the retrieval slowing (caused by extraposition/extraction), a prediction of underadditivity ensues because the postponement due to the bottleneck is effectively hiding or obscuring the retrieval effects. Similar predictions would hold if some other stage besides retrieval were affected by

Staum Casasanto & Sag (in press), some grammatical constraint violations combine additively and others under-additively in judgment tasks (see also Chapman 1974; Keller 2000; Sorace & Keller 2005). In short, additive and under-additive patterns are quite possible when grammar is in the mix. A super-additive pattern of results — where the net penalty is greater than one would expect on the basis of independent effects — is in principle consistent with a role for grammar, as well. As Sprouse, Wagers & Phillips (2012) point out, this pattern is ambiguous: it could either imply that two processing costs have combined such that their effects are magnified when they occur in tandem, or that some other factor — grammar — adds to the effects of two processing costs. Thus, had we found super-additive effects of combining extraposition and extraction, we would have been left with an ambiguous pattern. In such a case, further tests, methods, and comparisons between tasks would be necessary to resolve the source of unacceptability (see, for instance, Sprouse, Wagers & Phillips 2012; Hofmeister, Staum Casasanto & Sag in press).[8]

What is critical to the interpretation of how various factors combine is what is assumed about them in the first place. As noted earlier, there is no reason to ascribe the independent effects of extraction or extraposition to grammar; these effects fall under the purview of sentence processing theory and parsing principles. If we had considered some other acceptability phenomenon which involved two simultaneous grammatical violations and found that summing the effects of each independent grammatical violation equaled the effect of combining them, it would be illogical to posit some additional grammatical constraint that operated only when both violations were present. Thus, it is the analysis of the ingredients that constrains how we interpret a pattern of combination.

An outstanding challenge for any cost-driven account of judgment effects is to identify what these processing costs are, what the mechanisms are that produce the costs, and why they aggregate they way they do within a sentence. The processing costs of filler-gap dependencies have been well-documented in the literature (Wanner & Maratsos 1978; Kluender & Kutas 1993a; King & Kutas 1995). It thus seems uncontroversial to attribute the acceptability-lowering effects of extraction to processing costs linked to locating the appropriate integration site and retrieval difficulty. Here, we have assumed that extraposition also lowers judgments due to the gradient processing costs of linking an extraposed constituent with its syntactic host. But unlike extraction costs, which are strongly tied to retrieval difficulty, the precise reason why extraposition increases processing difficulty is less well-understood. Creating discontinuous constituency may, in principle, give rise to processing difficulty because of retrieval difficulty, a high degree of surprise, or the need to revisit a prior syntactic analysis. As noted previously, our experiment here was not designed to

---

extraction and extraposition, so long as this stage preceded the bottleneck (for further details, consult Pashler 1984; Pashler & Johnston 1989; Pashler 1994).

[8]Hofmeister, Staum Casasanto & Sag (in press) argue that only combinations of multiple processing costs yield super-additive interactions in judgment tasks, after testing various combinations of factors that lower acceptability. This conclusion, however, is not at all inconsistent with interpretations of super-additive interactions that reference grammar, as they note: "Even when we find super-additivity and/or a positive linear relationship between judgments and measures of memory/verbal abilities, these findings cannot definitively rule out the possibility that grammar or other factors account for some observable part of the variation in judgments. Given our current state of knowledge, it appears to be impossible to rule out grammar as a contributor to an acceptability contrast." Put slightly differently, Hofmeister et al. do not argue that super-additive interactions entail that the judgment pattern must be entirely explained in terms of sentence processing costs, but rather that two independent grammatical violations will not super-additively interact when co-present.

tease apart such theories. Nonetheless, we offer below some tentative views on the source of difficulty in sentences with extraposed constituents, based largely on the findings and discussion in Levy et al. (2012).

While distance clearly has a role to play in dependency processing due to memory retrieval costs (Gibson 1998; Gibson 2000; Bartek et al. 2011), it is less obvious why the relatively short distance between the critical NP and the PP, as in (13) below, should create serious retrieval difficulty:

(13)    Kenneth revealed that he overheard a nasty remark earlier about the President.

The distance between the NP *a nasty remark* and *about the President* here is minimal by just about any reckoning. The standard metric of distance in Gibson's Distance Locality Theory (DLT), for instance, would not predict that integrating the NP and PP would be any harder in the presence of an intervening modifier like *earlier* than in its absence. This is because DLT does not ascribe any cost to the intervening modifier because it does not introduce a discourse entity, i.e. it does not introduce a discourse-new individual or event referent.[9] Moreover, it is unlikely that the modifier has the potential to interfere substantially with the retrieval of the NP, given their general dissimilarity.

On a structural reanalysis account, in contrast, extraposition places demands on processing resources because (a) the adverb following the NP suggests that the NP representation is complete, i.e. the parser commits to an analysis, and (b) the presence of the PP forces the parser to revisit the original analysis. If the parser is incorrectly led to believe that the NP is finished or syntactically complete, then this belief grows stronger as more and more input argues in this direction (see discussion of 'digging in' in Tabor & Hutchins 2004 and effects of length of ambiguous region on recovery from garden path misanalysis in Ferreira & Henderson 1991). In essence, extraposed constituents are more difficult to process because they are more unexpected. Supporting this interpretation, Levy et al. (2012) found that the online difficulty associated with extraposed relative clauses varied systematically with structural expectations: strong expectations for a modifying relative clause facilitates comprehension of an extraposed relative clause. In cases of freezing, the memory costs of *wh*-dependency processing compound the reanalysis process as the point of retrieval and integration coincides with the reanalysis point. Given these observations, we suggest that structural reanalysis and expectation costs better account for the difficulty of extraposition, but we leave it for future research to corroborate or disprove this interpretation.

Whatever the preferred psycholinguistic account of both extraposition and extraction is, we have demonstrated that their independent effects on judgments are sufficient to account for the net unacceptability of 'freezing violations'. By obviating the need for a grammatical constraint to explain the unacceptability of such items, we ultimately arrive at a more streamlined grammar — one where certain structural configurations are technically licensed or generated by the grammar, but rarely if ever uttered due to the processing complexities involved, and when seen or heard in acceptability tasks are viewed as relatively unacceptable for the same reason. This has the advantage that grammars ultimately become simpler (Bever 1970; Bever 2009; Culicover & Jackendoff 2005; Hofmeister & Sag 2010;

---

[9]An alternative notion of 'distance' is discussed by Hawkins (1994, 1999), who calculates distance in terms of node quantity. As in our case, Hawkins remains non-committal about what processing mechanisms make longer filler-gap dependencies more complex, observing that the number of syntactic nodes is 'just one index of the relative complexity of an FGD, but it is a fundamental one with numerous correlating properties that involve additional processing operations.'

Featherston 2011). Yet another advantage of such an account is that it deals directly with not only generalizations about syntactic phenomena, but also exceptions to those generalizations. That is, a proposal like the Frozen Structure Constraint captures the general unacceptability of extraction from extraposition. What it does not do, however, is capture apparent counterexamples — cases where extraction from extraposition sounds relatively acceptable. Often times, syntactic accounts must treat such troublesome data as "peripheral" or due to secondary syntactic principles. In contrast, the relative acceptability of some sentences with extraction from extraposition can be dealt with straightforwardly on an account such as ours: we make the strong prediction that (judgments for) these sentences arise when various factors conspire to make the structure easier to process, e.g. supportive context, complex *wh*-phrases, highly predictable extraposed constituents, etc. A substantial advantage of a processing-based approach to freezing effects, therefore, is that we can view syntactic generalizations and variance surrounding those generalizations through the same mechanistic lens.

# 4 Appendix: Materials

1. Tell me which actor your friend read a story (about) twice (about).
You told me your friend read a story (about an actor) twice (about an actor).
2. I forgot which king Anna had started the play (about) repeatedly (about).
I forgot that Anna had started the play (about a king) repeatedly (about a king).
3. She wants to know which dictator you hung up a picture (of) last week (of).
She wants to know why you hung up a picture (of a dictator) last week (of a dictator).
4. We were told which thief the detective should be able to track down the report (about) easily (about).
We were told that the detective should be able to track down the report (about the thief) easily (about the thief).
5. I recalled which woman the musician sang a song (about) tearfully (about).
I recalled that the musician sang a song (about a woman) tearfully (about a woman).
6. I found out which music William had promised to file a complaint (about) as soon as possible (about).
I found that William had promised to file a complaint (about the music) as soon as possible (about the music).
7. Kenneth revealed which President he overheard a nasty remark (about) earlier (about).
Kenneth revealed that he overheard a nasty remark (about the President) earlier (about the President).
8. We found out which castle ruins archaeologists have searched for evidence (of) unsuccessfully (of).
We found that archaeologists have searched for evidence (of the castle ruins) unsuccessfully (of the castle ruins).
9. Frederick revealed which ballplayer he sold a signed picture (of) cheaply (of).
Frederick revealed that he sold a signed picture (of a ballplayer) cheaply (of a ballplayer).
10. The videotape showed which cards the dealer arranged the order (of) deftly (of).
The videotape showed that the dealer arranged the order (of the cards) deftly (of the cards).
11. I remembered which event Sally had mentioned plans (for) to the neighbors (for).
I remembered that Sally had mentioned plans (for the event) to the neighbors (for the

event).

12. It's perfectly clear which of the businesses Charlie stored the receipts (from) carefully (from).

It's perfectly clear that Charlie stored the receipts (from all the businesses) carefully (from all the businesses).

13. It is clear which burglary Carl filed a summary (of) with the police (of).

It is clear that Carl filed a summary (of the burglary) with the police (of the burglary).

14. Jan clarified which Greek god Susan found the sculpture (of) two days ago (of).

Jan clarified that Susan found the sculpture (of a Greek god) two days ago (of a Greek god).

15. The kids know which grandparents their mom wrapped gifts (from) at night (from).

The kids know that their mom wrapped gifts (from their grandparents) at night (from their grandparents).

16. I discovered which meeting they had to explain the purpose (of) very patiently (of).

I discovered that they had to explain the purpose (of the meeting) very patiently (of the meeting).

17. Fortunately, we were able to determine which experiment Simon had recorded the results (from) previously (from).

Fortunately, we were able to determine that Simon had recorded the results (from the experiment) previously (from the experiment).

18. Everyone could see which victims Marianne had rearranged snapshots (of) systematically (of).

Everyone could see that Marianne had rearranged snapshots (of the victims) systematically (of the victims).

19. Ted told me which monastery they had a good view (of) finally (of).

Ted told me that they had a good view (of the monastery) finally (of the monastery).

20. It was obvious which winnings Tom would not be able to save his share (of) for long (of).

It was obvious that Tom would not be able to save his share (of the winnings) for long (of the winnings).

# References

Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

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

Bartek, B., R. Lewis, S. Vasishth & M. Smith. 2011. In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37:1178–1198.

Bever, T. 1970. The cognitive basis for linguistic structures. In *Cognition and the Development of Language*, ed. J. R. Hayes, 279–362. New York: John Wiley & Sons.

Bever, T. 2009. Remarks on the individual basis for linguistic structures. In *Of minds and language: A dialogue with Noam Chomsky in the Basque country*, ed. M. Piattelli-Palmarini, J. Uriagereka & P. Salaburu, 279–295. Oxford: Oxford University Press.

Bever, T., J. Carroll & R. Hurtig. 1976. Analogy or ungrammatical sequences that are utterable and comprehensible are the origins of new grammars in language acquisition and linguistic evolution. In *An integrated theory of linguistic ability*, 149–82. NY: Crowell.

Chapman, R. 1974. *The interpretation of deviant sentences in English: A transformational approach.* The Hague: Mouton.

Chaves, R. P. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics* 49:285–327.

Chomsky, N. 1964. *Current issues in linguistic theory.* The Hague: Mouton de Gruyter.

Chomsky, N. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. S. Anderson & P. Kiparsky, 232–286. New York: Holt, Reinhart & Winston.

Chomsky, N. 1977. On *wh*-movement. In *Formal syntax*, ed. P. Culicover, T. Wasow & A. Akmajian, 71–132. New York: Academic Press.

Chomsky, N. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. 1986. *Barriers.* Cambridge: MIT Press.

Chomsky, N. 1995. *The Minimalist program.* Cambridge: MIT Press.

Christensen, K. R., J. Kizach & A. M. Nyvad. 2013. Escape from the island: Grammaticality and (reduced) acceptability of *wh*-island violations in danish. *Journal of Psycholinguistic Research* 42:51–70.

Clark, H. H. & S. E. Haviland. 1974. Psychological processes as linguistic explanation. In *Explaining linguistic phenomena*, ed. D. Cohen, 91–124. Washington, D.C.: Hemisphere.

Corver, N. 2006. Freezing effects. In *The Blackwell companion to syntax*, ed. M. Everaert & H. van Riemsdijk, 383–406. Malden, MA: Blackwell Publishing.

Cowart, W. 1997. *Experimental syntax: Applying objective methods in sentence judgments.* Thousand Oaks, CA: Sage Publications.

Culicover, P. & R. Jackendoff. 2005. *Simpler syntax.* Oxford University Press, USA.

Culicover, P. & S. Winkler. 2012. Freezing: A conspiracy. Submitted.

Deane, P. 1991. Limits to attention: a cognitive theory of island phenomena. *Cognitive Linguistics* 2:1–63.

Fanselow, G. & S. Frisch. 2006. Effects of processing difficulty on judgments of acceptability. In *Gradience in Grammar*, ed. G. Fanselow, C. Fery, M. Schlesewsky & R. Vogel, 291–316. Oxford: Oxford University Press.

Featherston, S. 2011. Three types of exceptions – and all of them rule-based. In *Expecting the unexpected: exceptions in grammar*, ed. H. S. . H. Wiese, 291–324. Berlin: Walter de Gruyter.

Ferreira, F. & J. Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 30:725–745.

Gelman, A. & J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gibson, E. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68:1–76.

Gibson, E. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In *Image, language, brain*, ed. A. Marantz, Y. Miyashita & W. O'Neil, 95–126. Cambridge, MA: MIT Press.

Gibson, E., S. Piantadosi & K. Fedorenko. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5:509–524.

Goodluck, H. & E. Zweig. 2013. Introduction: Formal vs. processing explanations of linguistic phenomena. *Language and Cognitive Processes* 28:1–8.

Grodner, D. & E. Gibson. 2005. Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29:261–290.

Hawkins, J. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75:244–285.

Hawkins, J. 2004. *Efficiency and complexity in grammars*. Cambridge: Oxford University Press.

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

Hofmeister, P., T. F. Jaeger, I. Arnon, I. Sag & N. Snider. 2013. The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes* 48:48–87.

Hofmeister, P., T. F. Jaeger, I. A. Sag, I. Arnon & N. Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in search of its evidential base*, ed. S. Featherston & W. Sternefeld, 185–206. Berlin: Mouton de Gruyter.

Hofmeister, P. & I. Sag. 2010. Cognitive constraints and island effects. *Language* 86:366–415.

Hofmeister, P., L. Staum Casasanto & I. A. Sag. 2013. Islands in the grammar? Standards of evidence. In *Experimental syntax and the islands debate*, ed. J. Sprouse & N. Hornstein, 42–63. Cambridge: Cambridge University Press.

Hofmeister, P., L. Staum Casasanto & I. A. Sag. in press. Processing effects in linguistic judgment data: (Super-)additivity and reading span scores. *Language and Cognition* to appear, 2014.

Huck, G. & Y. Na. 1990. Extraposition and focus. *Language* 66:51–77.

Jackendoff, R. & P. Culicover. 1972. A reconsideration of dative movement. *Foundations of Language* 7:397–412.

Just, M. & P. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99:122–149.

Keller, F. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.D. thesis, University of Edinburgh. College of Science and Engineering. School of Informatics.

Kim, I. 2013. Rethinking "island effects" in Korean relativization. *Language Sciences* 38:59–82.

King, J. & M. Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30:580–602.

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

Kluender, R. 1991. *Cognitive constraints on variables in syntax*. Ph.D. thesis, University of California, San Diego, La Jolla, CA.

Kluender, R. 1992. Deriving islands constraints from principles of predication. In *Island constraints: Theory, acquisition and processing*, ed. H. Goodluck & M. Rochemont, 223–258. Dordrecht: Kluwer.

Kluender, R. 1998. On the distinction between strong and weak islands: a processing perspective. In *Syntax and semantics 29: The limits of syntax*, ed. P. Culicover & L. McNally, 241–279. San Diego, CA: Academic Press.

Kluender, R. 2004. Are subject islands subject to a processing account? In *WCCFL 23 Proceedings*, ed. B. Schmeiser, V. Chand, A. Kelleher & A. Rodriguez. Somerville, MA: Cascadilla Press.

Kluender, R. 2005. Are subject islands subject to a processing account? In *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, ed. V. Chand, A. Kelleher, A. J. Rodríguez & B. Schmeiser, 475–499. Somverville, MA: Cascadilla Press.

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

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

Levy, R., E. Fedorenko, M. Breen & E. Gibson. 2012. The processing of extraposed structures in English. *Cognition* 12:12–36.

Marks, L. 1968. Scaling of grammaticalness of self-embedded English sentences. *Journal of Verbal Learning and Verbal Behavior* 7:965–967.

Miller, G. & N. Chomsky. 1963. Finitary models of language users. In *Handbook of mathematical psychology*, volume 2, ed. R. D. Luce, R. R. Bush & E. Galanter, 419–492. New York: Wiley.

Müller, S. 2004. Complex NPs, subjacency, and extraposition. *Snippets* 8:10–11.

Pashler, H. 1984. Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance* 10:358–377.

Pashler, H. 1994. Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin* 116:220–244.

Pashler, H. & J. Johnston. 1998. Attentional limitations in dual-task performance. *Attention* 155–189.

Pashler, H. & J. C. Johnston. 1989. Chronometric evidence for central postponement in temporally overlapping tasks. *The Quarterly Journal of Experimental Psychology* 41:19–45.

Phillips, C. 2006. The real-time status of island phenomena. *Language* 82:795–823.

Phillips, C. 2013. Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28:156–187.

Pinheiro, J. C. & D. M. Bates. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pylyshyn, Z. 1973. The role of competence theories in cognitive psychology. *Journal of Psycholinguistic Research* 2:21–50.

Ross, J. R. 1967. *Constraints on variables in syntax*. Ph.D. thesis, MIT, Cambridge, MA.

Schütze, C. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Sorace, A. & F. Keller. 2005. Gradience in linguistic data. *Lingua* 115:1497–1524.

Sprouse, J. 2007. *A program for experimental syntax*. Ph.D. thesis, University of Maryland, College Park, College Park.

Sprouse, J. 2009. Revisiting satiation: evidence for an equalization response strategy. *Linguistic Inquiry* 40:329–341.

Sprouse, J. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43:155–167.

Sprouse, J. & D. Almeida. to appear. The role of experimental syntax in an integrated cognitive science of language. In *The Cambridge handbook of biolinguistics*, ed. K. Grohmann & C. Boeckx. Cambridge: Cambridge University Press.

Sprouse, J., S. Fukuda, H. Ono & R. Kluender. 2011. Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax* 14:179–203.

Sprouse, J., M. Wagers & C. Phillips. 2012. A test of the relation between working memory and syntactic island effects. *Language* 88:82–123.

Staum Casasanto, L., P. Hofmeister & I. A. Sag. 2010. Understanding acceptability judgments: Additivity and working memory effects. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, ed. S. Ohlsson & R. Catrambone, 224–229. Austin, TX: Cognitive Science Society.

Tabor, W. & S. Hutchins. 2004. Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:431–450.

Wanner, E. & M. Maratsos. 1978. An ATN approach to comprehension. In *Linguistic theory and psychological reality*, ed. M. Halle, J. Bresnan & G. A. Miller, 119–161. Cambridge, MA: MIT Press.

Warren, T. & E. Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85:79–112.

Warren, T. & E. Gibson. 2005. Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes* 20:751–767.

Wexler, K. & P. Culicover. 1980. *Formal principles of language acquisition.* Cambridge, MA: MIT Press.