# Processing Factors Influencing Acceptability in Extractions from Complex Subjects

David R. Clausen

June 7, 2010

## 1 Introduction

This paper explores the processing factors influencing the acceptability of fillergap dependencies where the gap is located within a Complex Subject NP. These environments have traditionally been considered strong syntactic islands which prohibit hosting a gap. Contrary to theories which posit a categorical prohibition on gaps within certain syntactic island environments and account for this in terms of structural constraints in the competence grammar, we find systematic gradience in the acceptability of gaps located within Complex Subjects. We argue that a processing based acount, where acceptability is graded and determined by processing complexity, gives a more nuanced account of the data. We conduct a series of acceptability judgement experiments that systematically manipulate factors known to influence processing difficulty and find effects within the domain of Complex Subject NPs. We find that gaps located within Complex Subjects, while significantly less acceptable than gaps located in sentences without Complex Subjects, are nowhere near completely ungrammatical. Furthermore their acceptability can be systematically increased through the reduction of other processing difficulties. We also find that the acceptability of certain sentences that do not violate island constraints has been systematically overstated. By using acceptability judgements we reveal a nuanced view of island constraints that challenges the binary grammaticality judgments used to theorize about Island Constraints and argues for a processing based account.

## 2 Islands Constraints and Processing

## 2.1 Island Constraints Overview

Since Ross (1967), linguists have recognized certain syntactic environments, known as syntactic islands, which do not readily host gaps. Ensuing research focused on crosslinguistic variation in which syntactic environments disprefer gaps. Other work concentrated on identifying mitigating factors which allow a syntactic island to host a gap when it otherwise would not (Engdahl 1983). Most research focused on idenfiying the structural factors shared across syntactic islands in an effort to construct a unified account of island phenomena (Chomsky 1973). Most early work assumed syntactic islands should be accounted for as structural constraints on competence grammar. More recent work has questioned the categorical nature of syntactic islands (Kluender 1991, 1992, 1993a, 1993b, 1998, 2005). These observations drove increased exploration of syntactic islands as a processing phenomena (Hofmeister 2007). The remainder of this section will give an overview of the structural accounts of island phenomena and review the processing literature relevant to syntactic islands.

Ross (1967) categorized a wide variety of syntactic environments which do not host gaps. For the purposes of this paper we will focus mainly on Complex Subject islands illustrated in (1) and to a small extent on Factive islands illustrated in (2).

(1) \*Which book did  $[_{IP}]_{NP}$  that Sandy read GAP surprise Kim]?

(2) \*Why did  $[_{IP}$  they realize  $[_{IP}$  they could help them GAP]?

### 2.2 Structural Accounts of Island Constraints

Within the generative syntax tradition, island phenomena were handled by introducing constraints on transformations. The first major unified attempt to capture the effect of island constraints on the competence grammar was put foward in Chomsky (1973). Chomsky introduces the principle of Subjacency which restricts transformations between certain syntactic structures.

#### Subjacency:

No rule may move a phrase from position Y to position X in:

 $\ldots \, X \, \ldots \, \left[_{\alpha} \, \ldots \, \left[_{\beta} \, \ldots \, Y \, \ldots \, \right] \, \ldots \, \right] \, \ldots \, X \, \ldots \,$ 

Where  $\alpha$  and  $\beta$  are cyclic nodes.

In English, cyclic nodes are IP and NP. Subjacency represent a restriction on the number of nodes of a particular type an unbounded dependency may cross. Both Complex Subject islands and Factive islands are categorically ruled out by Subjacency.

The later Government and Binding approach replaced the notion of Subjacency with the notion of "barriers" to movement (Chomsky 1986). The somewhat more complex definition is given below.

#### **Barriers:**

 $\gamma$  is a Blocking Category(BC) for  $\beta$  if and only if  $\gamma$  is not L-marked and  $\gamma$  dominates  $\beta.$ 

 $\gamma$  is a barrier for  $\beta$  if and only if (a) or (b):

a.  $\gamma$  immediately dominates  $\delta$  where  $\delta$  is a BC for  $\beta$ ,

b.  $\gamma$  is a BC for  $\beta$ ,  $\gamma \neq$  IP.

Again this definition rules out both Complex Subject islands and Factive islands. Key to the "barriers" approach is a multi-valued notion of Subjacency violation where the more "barriers" between a filler and gap, the greater the violation and the lower the acceptability.

Outside of the transformational syntax tradition, accounts of island phenomena make use of feature propagation along with general constraints on the relation of features between mothers and daughters. An early account by Gazdar et al. (1985) outlines principles adopted in later frameworks.

#### Features:

Fillers are related to gaps through the [SLASH] feature.

[SLASH] appears on all nodes on the path between the filler and the gap.

When [SLASH] appears on a daughter it must appear on the mother.

When [SLASH] appears on the mother it must appear on the head.

The approach of Gazdar et al. rules out Complex Subject islands which would contain a configuration where the [SLASH] feature appears on the VP but not on the Verb head. Like the Subjacency account, the feature propagation approach predicts a categorical distinction between gaps located within a syntactic island and gaps located elsewhere.

A small but growing literature has challenged the categorical nature of island constraint violations by suggesting non-structural factors which influence their acceptability. Kluender (2005) identifies a number of factors which influence the acceptability of Complex Subject island violations based on intuitive judgements. These factors include the type of NP subject, the presence of discourse referents between the filler and the gap, and the position of the gap within the complex subject. Examples (3)-(5) illustrate an increase in acceptability as the location of the gap and the type of NP subject are manipulated, with (3) the least acceptable and (5) the most acceptable.

(3) Who does [that she can bake ginger cookies for *GAP*] give her great pleasure?

(4) What does [baking GAP for your grandchildren] tire you out? <

(5) Who does [baking ginger cookies for GAP] tire you out?

In light of these and other examples, recent research has focused on developing accounts of syntactic islands that allow for the graded nature of their acceptability (Hofmeister & Sag 2010). Central to these accounts is the notion that some features of island phenomena are not driven by properties of the competence grammar, but instead are the result of general processing constraints. We now turn to review some of the research on the relationship between processing and syntactic islands.

## 2.3 Processing Syntactic Islands

Central to a processing account of syntactic islands is a basic conception of how the parser handles filler-gap dependencies. It is generally accepted that the parser maintains the filler in memory and must retrieve it at the correct gap site. Because gaps are not vocalized and can usually only be conclusively located when the parser encounters material following the gap, the process of actively locating the gap site introduces substantial processing complexity. To minimize overall processing complexity one hypothesis is that the parser often posits a gap at the earliest possible location, giving rise to the Active Filler Hypothesis (AFH) (Frazier & Flores D'Arcais 1989).

#### Active Filler Hypothesis:

The Parser attempts to form a filler-gap dependency at the earliest opportunity.

The AFH is supported by evidence from the "filled gap effect", where we find increased processing time when the parser encounters an NP in an argument position that could have been a gap. We also find evidence from the "semantic plausibility effect", where we see effects of the semantic match between a filler and a verb that could potentially host the gap. The parser experiences increased processing effort at a verb that could potentially head a gap when the filler is semantically appropriate indicating that the parser is positing a gap at that location. Little evidence for either the "filled gap effect" or the "semantic plausability effect" has been found within syntactic island environments. This finding suggests that the parser is sensetive to structural prohibitions on the locations of gaps, giving early support to a competence based account of island phenomena.

More recent work has taken two forms. The first takes as its starting point the categorical nature of syntactic islands and concentrates on evaluating whether the parser posits gaps within island environments in order to answer the question of whether the parser is sensitive to the competence grammar (Philips 2006, Wagers 2009). This work finds that the parser is generally sensitive to the Coordinate Structure Constraint. They find mixed results when evaluating how the parser handles parasitic gaps based on Subject and Adjunct islands. The second has taken as its basis the observations that grammatical acceptability is a graded phenomena and explores the processing factors which influence processing difficulty and the resulting acceptability associated with syntactic islands (Hofmesiter 2007, Hofmeister & Sag 2010). This work has found processing factors, such as the complexity of the filler, which mitigate processing difficulty and systematicly increase acceptability across a variety of island constraint violations including Adjunct islands, Complex NP islands and Wh-islands.

The remainder of this paper will focus on evaluating the hypothesis put forward in Hofmeister (2007) that the unacceptability of island constraint violations is due to their increased processing complexity. We will look specifically at island constraint violations where the gap is hosted within a complex subject NP. This work relies on the link between processing complexity and acceptability, where increased processing complexity leads to decreased acceptability, that has been robustly confirmed across the processing literature. In three sets of experiments we systematically manipulate factors known to increase processing complexity, such as reducing filler informativity or discourse referent accessibility, and expect to see decreased acceptability. We also test the observations made in Kluender (2005) regarding the location of the gap within the complex subject NP by manipulating the main verb in the complex subject.

## **3** Complex Subjects and Extraction

### 3.1 Overview

This set of experiments examines properties of complex subjects. Using acceptability judgments, we compare the effect of extraction from complex subjects to the control case of declarative sentences containing complex subjects without extraction. Additionally, we compare extraction from complex subjects to extraction from complement complex NPs, which are semantically similar except for taking an experiencer as a subject as opposed to an object.

These experiments make use of complex NPs taking the form of matched gerundive verb-noun pairs. We selected gerundive verb noun pairs by examining a large newswire corpus and computing the Pointwise Mutual Information between the verbs and their direct objects. Pointwise Mutual Information between two words A and B, is defined in the following equation:  $PMI(A,B) = \log_2(\frac{P(A,B)}{P(A) \times P(B)})$ . This is the log base 2 of the probability that the two words occur together divided by the product of their individual probabilities. This represents how much uncertainty about B decreases when A is known or how far from independence the two distributions are. Intuitively, two words with high Pointwise Mutual Information occur together very frequently and almost never appear independently, while two words with low Pointwise Mutual Information occur together rarely and occur independently often. In our case, A represents the verb and B its direct object. Their Pointwise Mutual Information can be thought of as the amount we know about the identity of the direct object when we only know the identity of the verb. Pointwise Mutual Information is a symmetric relationship, that is PMI(A,B) = PMI(B,A). This is important because in the case of wh-extraction we want to know how much knowing the filler tells us about the verb it is in an unbounded dependency with. filler-verb pairs with high Pointwise Mutual Information should bias the parser to posit a gap as one of the arguments of the verb. In a pilot study we randomly selected 24 verb-noun pairs ranging form 1.29 bits of Pointwise Mutual Information (analyzing, piece) to 9.69 bits (itemizing, deduction) and had subjects rate their acceptability when used as the filler and the gerundive verb in a complex subject question. Pointwise Mutual Information was computed over verb-direct object pairs taken from the AP Newswire and New York Times sections of the English Gigaword Corpus (Graf and Cierci 2003). An example question is given in (6). An interesting non-linear relationship between acceptability and Pointwise Mutual Information was found, although we will not discuss this finding further. We found acceptability peaked at 4.5 bits of mutual information and selected matched verb-noun pairs from a range of 4.4 to 4.6 bits of mutual information for the following experiments.

(6) What piece will analyzing be easy?

All experiments were conducted using WebSPR, a set of software tools that facilitate online psycholinguistic experiments. Subjects were recruited and managed through Amazon Mechanical Turk, an online marketplace for short human activities. Subjects were recruited to "Rate Sentence Acceptability" and compensated for their involvement. After giving consent, subjects were directed to the experiment website and instructed to "Please rate the acceptability of the following sentences on a scale from 1 (Completely unacceptable) to 7 (Completely Acceptable)." They were then presented with each sentence one at a time in the center of the screen above seven boxes numbered 1 (Completely Unacceptable) to 7 (Completely Acceptable); they could click on a box or enter an appropriate number on their keyboard to give their rating. Recent work has demonstrated the experimental validity of conducting psycholinguistic research over Amazon Mechanica Turk (Munro et al. 2010).

### 3.2 Experiment 1: Declarative Complex Subjects

#### 3.2.1 Design

Experiment 1 examines the acceptability of declarative sentences containing complex subjects in order to establish a baseline for further comparison. This experiment is a 2x3 design with two levels manipulating the direct object of the complex subject and the matrix verb (SUBJ and OBJ) and three levels manipulating the subject of the complex subject (POSS, NULL and NAME). Each participant rated the acceptability of 20 randomly ordered sentences including 12 target sentences and 8 filler sentences. All filler sentences were basic wh-question sentences without island constraint violations. To control for differences across experimental conditions we used a set of 12 templates to provide a common sentence frame across conditions. Each target sentence was constructed from one sentence template as described below.

#### 3.2.2 Materials

72 stimuli were produced from 12 templates by manipulating the 6 experimental conditions. On average it took six subjects to see the entire set of stimuli. Templates consisted of a complex gerundive NP, a template noun matched by Pointwise Mutual Information, a proper name and either the matrix verb *bother* or *upset*. We then constructed declarative sentences by selecting between a possessive subject my (POSS), a null subject (NULL), or a distinct proper name (NAME) for the gerundive NP. Finally, we selected between two configurations, one (SUBJ), where the template noun is the direct object of the complex gerundive NP and the template proper name is the matrix verb direct object, and another where their positions are exchanged (OBJ). The (SUBJ) condition has high Pointwise Mutual Information between the head of the complex NP and its direct object, while the (OBJ) condition does not. Examples of all 6 conditions for one template follow. A full stimulus list can be found in Appendix A.

Cond	Subj	Verb	Noun	Matrix	Noun
SUBJxPOSS	My	disciplining	a student	would upset	Chris.
SUBJxNULL		Disciplining	a student	would upset	Chris.
SUBJxNAME	John	disciplining	a student	would upset	Chris.
OBJxPOSS	My	disciplining	Chris	would upset	a student.
OBJxNULL		Disciplining	Chris	would upset	a student.
OBJxNAME	John	disciplining	Chris	would upset	a student.

Template: disciplining, student, Chris, upset.

#### 3.2.3 Participants

90 participants who declared themselves native speakers of English and connected from an IP address located in the United States completed the experiment in an average of 182 seconds each. While this may appear fast, each subject spent on average 9 seconds reading and judging each sentence which provides enough time to read the sentence and briefly reflect on its acceptability. Most subjects entered their ratings using the number pad, further speeding the rating process. Nearly all subjects participated in only one experiment, although a small percentage of users were subjects in multiple experiments. No participant was a subject in the same experiment multiple times.

#### 3.2.4 Results

All experiments collected reaction time data for each rating, and we used this data to remove response outliers. We pooled responses across experiments and then performed the following data cleaning procedures. Responses that took more than 20,000ms and less than 1,000ms were excluded on the basis that subjects were not following directions to read the sentences naturally and rate them promptly. We further excluded responses that had reaction times that were more than 3 standard deviations from the mean response time. These procedures reduced the total number of sentence ratings for all experiments from 10,860 to 9617.

In Experiment 1 participants rated target sentences (Mean = 4.96, SD = 1.81) significantly lower than filler sentences (Mean = 6.02, SD = 1.68, p < .05). Participants also rated the SUBJ conditions (Mean = 5.163, SD = 1.823) significantly higher than the OBJ conditions (Mean = 4.763, SD = 1.772, p < .05). This result is likely due to the high Pointwise Mutual Information between discipline and a student and the low Pointwise Mutual Information between discipline and Chris. A discussion of the results for the three gerund subject

conditions for all experiments is deferred until section 6, where we discuss the effects of discourse referents on the processing of complex subjects.

## 3.3 Experiment 2: Extraction From Complex Subjects

#### 3.3.1 Design

Experiment 2 examines the effect of *wh*-extraction from complex subjects on acceptability ratings. We compare extraction from the direct object of a gerundive complex subject to extraction of the direct object of the matrix verb. The experiment is a 2x3 design with two levels manipulating the site of extraction, either the direct object of the complex subject or the direct object of the matrix verb (SUBJ and OBJ), and three levels manipulating the subject of the complex subject (POSS, NULL and NAME). Each subject rated the acceptability of 20 randomly ordered sentences including 12 target sentences and 8 filler sentences. The 12 target sentences contained one sentence built from each template and two sentences from each condition as described below and were randomized across subjects.

### 3.3.2 Materials

72 stimuli were produced from 12 templates by manipulating the six experimental conditions. Templates consisted of a complex gerundive NP, a matched noun, a proper name and either the matrix verb *bother* or *upset*. Target sentences were constructed by selecting the gap location, either the direct object of the complex gerundive subject (SUBJ) or the direct object of the matrix verb (OBJ). We then selected between a possessive subject my (POSS), a null subject (NULL), or a distinct proper name (NAME) for the gerundive NP. Examples of all six conditions for one template follow. A full stimulus list can be found in Appendix A.

Template: appointing, commissioner, Joe, bother.

SUBJxPOSS: Which commissioner would my appointing bother Joe?

**SUBJxNULL:** Which commissioner would appointing bother Joe?

SUBJxNAME: Which commissioner would Steve appointing bother Joe?

**OBJxPOSS:** Which commissioner would my appointing Joe bother?

**OBJxNULL:** Which commissioner would appointing Joe bother?

**OBJxNAME:** Which commissioner would Steve appointing Joe bother?

#### 3.3.3 Participants

120 participants who declared themselves native speakers of English and connected from an IP address located in the United States completed the experiment in an average of 217 seconds each.

#### 3.3.4 Results

Participants rated all target sentences (Mean = 2.83, SD = 1.83) significantly lower than declarative sentences (Mean = 4.96, SD = 1.81, p < .05). Subjects rated extractions from complex subjects (Mean = 2.61, SD = 1.72) significantly lower than extractions from the matrix direct objects (Mean = 3.04, SD =1.90, p < .05).

## 3.4 Experiment 3: Extraction From Gerundive Complements

#### 3.4.1 Design

Experiment 3 examined extraction from gerundive complements of the verbs *regret* and *dislike*. Semantically they are similar to *upset* and *bother* verbs used in the other experiments in also being "psych-verbs". Syntactically, these verbs take an experiencer and gerundive NP complements which are considered factive islands. This experiment tests the distinction in acceptability ratings between the weak factive islands and the strong complex subject islands. The experiment is a 2x3 design with two levels manipulating the site of extraction, either subject of the matrix verb or the direct object of the gerundive complement (SUBJ, OBJ), and three levels manipulating the subject of the complex subject (POSS, NULL and NAME). Each subject rated the acceptability of 20 randomly ordered sentences including 12 target sentences and 8 filler sentences. The 12 target sentences contained one sentence built from each template and two sentences from each condition as described below and were randomized across subjects.

#### 3.4.2 Materials

72 stimuli were produced from 12 templates by manipulating the 6 experimental conditions. Templates consisted of a complex gerundive NP, a matched noun, a proper name and either the matrix verb *regret* or *dislike*. Target sentences were constructed by selecting the gap location, either the entire matrix subject (OBJ) or as the direct object of the gerundive NP complement (SUBJ). We then selected between a possessive subject my (POSS), a null subject (NULL), or a distinct proper name (NAME). Examples of all six conditions for one template follow. A full stimulus list can be found in Appendix A.

Template: intimidating, scientist, Greg, dislike.

SUBJxPOSS: Which scientist would Greg dislike my intimidating?

SUBJxNULL: Which scientist would Greg dislike intimidating?

SUBJxNAME: Which scientist would Greg dislike Kate intimidating?

**OBJxPOSS:** Which scientist would dislike my intimidating Greg?

**OBJxNULL:** Which scientist would dislike intimidating Greg?

**OBJxNAME:** Which scientist would dislike Kate intimidating Greg?

#### 3.4.3 Participants

90 participants who declared themselves native speakers of English and connected from an IP address located in the United States completed the experiment in an average of 205 seconds each.

### 3.4.4 Results

Overall participants rated target sentences (Mean = 4.84, SD = 1.84) significantly higher than target sentences in Experiment 2 (Mean = 2.83, SD = 1.83, p < .05). Participants rated extractions of the matrix subject (Mean = 4.71, SD = 1.87) slightly lower than extraction from the gerundive complement (Mean = 4.95, SD = 1.81, p < .05).

### 3.5 Discussion

These three experiments have established the basic properties of complex subjects pertinent to this paper. For an overview refer to Figure 1a, which provides a comparison of the distribution of ratings across the three experiments, Experiment 1 (control), Experiment 2 (mi) and Experiment 3 (admire) broken down by SUBJ (Blue) or OBJ (Green) conditions, and Figure 1b, which provides a similar perspective after z-score normalizing the ratings. Z-score normalizing is accomplished by subtracting from each participants ratings their mean rating and dividing by two times the standard deviation of the subject's ratings. In Experiment 1 and Experiment 3 the majority of the answer distribution is to the right of the scale in Figure 1a, while in Experiment 2 the distribution is heavily skewed to the low end of the scale, indicating the dramatic drop in acceptability for sentences with filler-gap dependencies and complex subjects. The distribution for both extractions from the complex subject and from the matrix direct object in Experiment 2 received low acceptability ratings even though extraction from the matrix direct object was slightly preferred as indicated by more of the distribution mass for the OBJ (Green) condition to the right of the SUBJ (Blue) condition. Similar patterns are evident in Figure 1b, which provides the same comparisons after transforming each rating by subtracting the participants mean rating and dividing by twice the standard deviation of the participants ratings. This visualization sacrifices absolute ratings differences between experiments but allows a direct comparison of the effect of SUBJ and OBJ conditions between experiments which will be valuable in later discussion.

Experiment 1 demonstrated that declarative sentences with complex subjects receive lower acceptability ratings than sentences without complex subjects even in the absence of any filler-gap dependencies. This finding comports well with numerous production and aquisition studies which indicate that complex subjects are significantly harder to process than other structures of similar syntactic complexity (Kluender 2005, Clark & Wasow 1998, Kemper 1987). This





evidence alone is not enough to establish the link between processing complexity and acceptability judgements, but in light of recent work linking the two it is highly suggestive (Hofmeister 2007, Hofmeister 2009).

Experiment 2 illustrated the primary empirical observation which gave rise to the the original Sentential Subject constraint, that extractions out of complex subjects lead to low acceptability if not outright ungrammaticality (Ross 1967). We see that questions containing gaps inside complex subjects are rated much lower on the acceptability scale than their simple declarative counterparts. It is surprising that sentences containing gaps in the direct object of the matrix verb are also rated significantly lower than their declarative counterparts although still higher than subject extractions. These sentences violate no known island constraint; instead, we appear to have an interaction between the complexity introduced by the filler-gap dependency and the complexity introduced by a complex subject. When both sources of processing complexity are present in the sentence we see a large reduction in acceptability. When the gap is found within the complex subject, this reduces acceptability further, but not nearly as much as the basic interaction between the presence of a complex subject and the presence of a filler-gap dependency. Although the low acceptability could be a result of poor stimulus design, the overall high acceptability of other stimuli constructed by similar methods in Experiments 1 and 3 indicate that this interpretation is unlikely. It appears that the theoretical literature has drastically overstated the difference in acceptability between extractions from complex subjects and extractions from other syntactic positions.

Experiment 3 provided a minimal comparison between extractions from complex gerundive subjects in experiment 2 and complex gerundive complements. We see that the gerundive NP in subject position drastically reduces acceptability when it participates in a filler gap-dependency. In a surprising finding, we observed a slight preference for extraction from the gerundive complement over extraction of the entire matrix subject. Complements of factive verbs like realize and dislike are considered weak islands and extractions from these cenvironments should be dispreferred compared to extractions of a non-complex subject. In a previous pilot study we demonstrated that the Pointwise Mutual Information between the filler and the verb head of a gap in a complex subject can significantly increase or decrease acceptability depending on the amount of mutual information. Since we selected stimuli based on Pointwise Mutual Information to maximize acceptability ratings it is likely that this effect is actually overcoming the dispreference for violating the Factive Island constraint since the mutual information biases the parser to an interpretation where the gap is present in the gerundive complement and away from an interpretation where it is the matrix subject. In the sentence when the noun *student* is used as a filler it is quite likely to appear as the direct object of *discipline* because they have high Pointwise Mutual Information and less likely to appear as the subject of dislike. We now turn to our next set of experiments, designed to further explore the role of mutual information and informativity on the acceptability of complex subjects.

## 4 Complex Subjects and Informativity

### 4.1 Overview

A growing body of empirical work suggests that many island constraints are modulated by factors that influence processing complexity (Hofmeister 2007, Hofmeister 2009, Philips 2008, Wagers & Philips 2009, Hofmeister & Sag 2010). Reducing processing complexity improves the acceptability of sentences that contain an island constraint violation. In his dissertation, Hofmeister demonstrates that filler informativity plays a crucial role in the processing complexity and the resulting acceptability judgements of sentences involving filler-gap dependencies. His hypothesis holds that processing filler-gap dependencies requires encoding the filler and then retrieving the filler from memory at the appropriate gap location. The informativity of the filler plays an important role in this process by influencing the number of cues available to access the filler from memory at retrieval time. The more informative a filler is the more cues are available for retrieval and the easier this retrieval process will be. If we can make the process sufficiently easy, we can overcome the processing complexity normally introduced by filler-gap dependencies when a gap is located within a syntactic island and increase the overall acceptability of such a sentence.

Hofmeister defines relative informativity as follows: An expression x1 is more informative than an expression x2 if the semantic and syntactic information encoded by x2 is a proper subset of the information encoded by x1. Based on this definition there is a clear hierarchy of filler complexity that we exploit in the following two experiments (who < which student < which student from Kentucky). The more complex the filler, the more informative it is under Hofmeister's definition. A bare filler like who contains only the information that it stands for something human, while which student adds information that the thing is also a student. The information contained in who is a subset of the information contained in which student, ranking who lower than which student in informativity. When we add the intersective modifier from Kentucky to the form which student from Kentucky, we further increase informativity. In general, the more semantically specific a filler, the more informative it will be.

We can draw a connection between Hofmeister's definition of informativity and the concept of information developed in Shannon (1948). The information of a string is the negative log probability of that string occurring. The lower the probability of the string, the higher the amount of information is contains. If we take Hofmeister's scale, we see that the informativity of a filler increases as the number of words in the filler increases. Information also generally increases as the number of words in a string increases because the conditional probability of a word given the previous words is almost never 1.0. While this reframing of informativity in terms of information content is intriguing, we will not pursue its development in this paper except to note that a full treatment would require definitions of the information content contributed by all levels of linguistic structure. We introduce the connection to motivate the use of another concept from Information Theory, Pointwise Mutual Information introduced in Section 3.1. In Experiment 2 we made use of the mutual information between the filler and a gerundive verb. In Experiment 4 we use bare wh-fillers which have the lowest information content and have no Pointwise Mutual Information between filler and gerundive verb. High Pointwise Mutual Information and filler informativity should facilitate processing by increasing the parser's ability to predict the gap location. The reduction in processing complexity should lead to increased acceptability. In Experiment 4 we use bare wh-fillers, removing the benefit from Pointwise Mutual Information and filler informativity. As a result acceptability should decline for extractions from complex subjects. In Experiment 5 we reintroduce the Pointwise Mutual Information between filler and gerundive verb and also increase the filler information content by adding to the filler a prepositional phrase to test if Hofmeister's predictions hold for complex subject extractions.

## 4.2 Experiment 4: Extractions with Low Informativity Fillers

#### 4.2.1 Design

Experiment 4 mirrors Experiment 2, but replaces fillers of the form *which* N with fillers of the form *who*. As mentioned, this reduces the informativity of the filler, which should reduce the number of cues available at retrieval, increasing processing complexity. It also reduces the mutual information between the filler and the gerundive verb. Thus, the parser cannot make use of the high Pointwise Mutual Information between the filler and the head of the gap to facilitate resolution of the unbounded dependency.

The experiment is a 2x3 design with two levels manipulating the site of extraction, either the direct object of the complex subject or the direct object of the matrix verb (SUBJ, OBJ) and three levels manipulating the subject of the complex subject (POSS, NULL and NAME). Each subject rated the acceptability of 20 randomly ordered sentences, including 12 target sentences and 8 filler sentences. The 12 target sentences contained one sentence built from each template, with two sentences from each condition, and were randomized across subjects (see below).

### 4.2.2 Materials

72 stimuli were produced from 12 templates by manipulating the six experimental conditions. Templates consisted of a complex gerundive NP, a proper name and either the matrix verb *bother* or *upset*. Target sentences were constructed by selecting the gap location, either the direct object of the complex gerundive subject (SUBJ) or the direct object of the matrix verb (OBJ). We then completed the complex subject by selecting between a possessive subject my (POSS), a null subject (NULL), or a distinct proper name (NAME) for the gerundive NP. Examples of all six conditions for one template follow. A full stimulus list can be found in Appendix A. Template: deceiving, Sarah, bother.

SUBJxPOSS: Who would my deceiving bother Sarah?

SUBJxNULL: Who would deceiving bother Sarah?

SUBJxNAME: Who would Megan deceiving bother Sarah?

**OBJxPOSS:** Who would my deceiving Sarah bother?

**OBJxNULL:** Who would deceiving Sarah bother?

**OBJxNAME:** Who would Megan deceiving Sarah bother?

#### 4.2.3 Participants

90 participants who declared themselves native speakers of English and connected from an IP address located in the United States completed the experiment in an average of 200 seconds each.

#### 4.2.4 Results

Participants rated target sentences (Mean = 3.01, SD = 2.02) higher than target sentences in Experiment 2 (Mean = 2.83, SD = 1.83, p < .05). Subject extractions (SUBJ) received much lower acceptability ratings (Mean =2.43, SD = 1.73) than object extractions (OBJ) (Mean = 3.60, SD = 2.11, p <.05). Extractions from matrix objects(Mean = 3.60, SD = 2.11, p < .05). were rated significantly higher than extractions from matrix objects in Experiment 2 (Mean = 3.04, SD = 1.90, p < .05). Extractions from complex subjects in this experiment (Mean = 2.43, SD = 1.73) appeared to be rated lower than extractions from complex subjects in Experiment 2 (Mean = 2.61, SD = 1.72, p < .1), although this was not statistically significant. We will review a more sophisticated analysis that allows a better understanding of the sources of variation in acceptability ratings in the discussion section 4.4.

## 4.3 Experiment 5: Extraction With High Informativity Fillers

#### 4.3.1 Design

By adding a prepositional phrase to each noun heading a filler we increase the informativity of the filler and restore the Pointwise Mutual Information between the filler and the gerundive verb in the complex subject. Experiment 5 again, mirrors Experiment 2, except each filler of the form *which* N is replaced with a filler of the form *which* N P N. A filler in Experiment 2 such as *which perpetrator* becomes *which perpetrator with a motive*.

The experiment is a 2x3 design with two levels manipulating the site of extraction, either the direct object of the complex subject or the direct object of the matrix verb (SUBJ, OBJ) and three levels manipulating the subject

of the complex subject (POSS, NULL and NAME). Each subject rated the acceptability of 20 randomly ordered sentences including 12 target sentences and 8 filler sentences. The 12 target sentences contained one sentence built from each template with two sentences from each condition and were randomized across subjects (see below).

#### 4.3.2 Materials

72 stimuli were produced from 12 templates by manipulating the six experimental conditions. Templates consisted of a complex gerundive NP, a paired noun, a prepositional phrase, a proper name and either the matrix verb *bother* or *upset*. Prepositional phrases were selected by choosing high frequency trigrams beginning with the paired noun and ending in a two word prepositional phrase. Target sentences were constructed by selecting the gap location, either the direct object of the complex gerundive subject (SUBJ) or the direct object of the matrix verb (OBJ). Next, we constructed a filler by composing the paired noun with the prepositional phrase. We then completed the complex subject by selecting between a possessive subject my (POSS), a null subject (NULL), or a unique proper name (NAME) for the gerundive NP. Examples of all six conditions for one template follow. A full stimulus list can be found in Appendix A.

**Template:** arresting, perpetrator, with a motive, Susan, bother.

- **SUBJxPOSS:** Which perpetrator with a motive would my arresting bother Susan?
- **SUBJxNULL:** Which perpetrator with a motive would arresting bother Susan?
- **SUBJxNAME:** Which perpetrator with a motive would Chris arresting bother Susan?
- **OBJxPOSS:** Which perpetrator with a motive would my arresting Susan bother?
- **OBJxNULL:** Which perpetrator with a motive would arresting Susan bother?
- **OBJxNAME:** Which perpetrator with a motive would Chris arresting Susan bother?

#### 4.3.3 Participants

90 participants who declared themselves native speakers of English and connected from an IP address located in the United States completed the experiment in an average of 200 seconds each.

#### 4.3.4 Results

Participants rated target sentences (Mean = 2.74, SD = 1.73) lower than target sentences for Experiment 4 (Mean = 3.01, SD = 2.02, p < .05) but not significantly differently from target sentences in Experiment 2 (Mean = 2.83, SD =1.83, p = .29). Participants did not rate complex subject (SUBJ) extractions (Mean = 2.67, SD = 1.65) significantly differently from matrix object (OBJ) extractions (Mean = 2.81, SD = 1.80, p = .20). Extractions from complex subjects and matrix objects were not rated significantly different from similar extractions in Experiment 2.

### 4.4 Discussion

Experiments 2, 4 and 5 explored the effect of filler informativity and filler-gap Pointwise Mutual Information on acceptability ratings for sentences containing extractions from complex subjects. For an overview refer to Figure 2a, which provides a comparison of the distribution of ratings across the three experiments Experiment 2 (mi), Experiment 4 (bare) and Experiment 5 (complex) broken down by SUBJ (Blue) or OBJ (Green) conditions and Figure 2b which provides a similar perspective after z-score normalizing the ratings. There is a significant bias towards low acceptability ratings in the complex subject extraction condition of Experiment 4 (bare). As the informativity of the filler increases and we restore the Pointwise Mutual Information between filler and gap in Experiment 2 (mi), we see the blue distribution representing the subject extraction condition become less skewed to the low end of the rating scale. There is also a slight increase in skew towards the negative end of the scale for the object extraction condition, plotted in Green. When the filler informativity increases further in Experiment 5, the two conditions merge into very similar distributions. After applying z-score transformations to the raw ratings, we see an even clearer representation of the effect of informativity on acceptability. As we increase filler informativity, the effect of decreased acceptability when extracting from the complex subject is mitigated and eventually disappears.

Experiment 4 demonstrated that by removing the mutual information between the filler and the gerundive verb in the complex subject while also reducing the informativity of the filler, we drastically reduce the acceptability of extractions from complex subjects compared to the results established in Experiment 2. We also observed an increase in acceptability for the matrix object extraction condition compared to the same condition in Experiment 2. While the lack of filler informativity explains the drop in acceptability for complex subject extractions in Experiment 4 compared to Experiment 2, it does not explain why this would increase the acceptability of the corresponding matrix object extractions. This suggests that the mechanism which resolves filler-gap dependencies makes positive and negative use of the Pointwise Mutual Information between a filler and a potential gap site head when positing a gap. If a head is identified with high Pointwise Mutual Information and a gap is indeed identified there, this will facilitate processing. If, on the other hand, the gap is found at a site with a





head that has low Pointwise Mutual Information with the filler, processing will be inhibited. We see this in a simultaneous increase in acceptability for complex subject extractions and a decrease in acceptability for matrix object extractions as we add Pointwise Mutual Information moving from Experiment 4 to Experiment 2. It is possible that the parser weighs relative probabilities of different potential gap sites, and it is this relative probability that ultimately influences processing complexity and the resulting acceptability judgements. Further work that carefully controls the Pointwise Mutual Information between the filler and all potential gap head sites is needed to explore this possibility.

Experiment 5 showed that further increasing the informativity of the filler by adding a prepositional phrase can mitigate the difference in acceptability ratings between extractions from the complex subject and from the matrix direct object. If the parser were truly sensitive to a grammatical prohibition on extraction from complex subjects, no manipulation of processing factors should remove the difference between the two extraction conditions. To the contrary, Experiment 2, 4 and 5 reveal a steady decrease in the difference between the two conditions as the informativity of the filler increases. While this difference is partially due to an increase in acceptability of the SUBJ condition between Experiment 4 and 5, it is much smaller than the decrease in acceptability of the OBJ condition from the low informativity Experiment 2 to medium informativity Experiment 4 to the high informativity Experiment 5. There is likely a more nuanced interaction between the Pointwise Mutual Information, filler informativity, and potential gap locations which deserves in-depth study.

Whatever mechanism allows the parser to resolve filler-gap dependencies, it relies on both filler informativity and Pointwise Mutual Information to perform this task. The low acceptability observed for extractions from complex subjects is also present in extractions from matrix direct objects. As we saw in Experiments 1 and 2, it is simply the presence of both a complex subject and a filler-gap dependency simpliciter that increases processing complexity and reduces acceptability regardless of the gap location. The slight further decrease in acceptability found when the gap is within a complex subject can be systematically mitigated by reducing the processing complexity associated with the filler-gap dependency by increasing informativity and using Pointwise Mutual Information to bias the parser or systematically increased by decreasing filler informativity and removing the Pointwise Mutual Information bias. The structural difference between having a gap within the complex subject or in the matrix direct object position is less important that the filler informativity in determining the acceptability of these sentences.

To confine the findings in Experiments 2, 4 and 5, we fit a Linear Mixed Effects model including experiment as a fixed effect and allowing the slope and intercept of the subject extraction SUBJ condition to vary across experiments. This allows us to examine how the different experiments modulate the difference in acceptability between extractions from complex subjects and extractions from the matrix direct object while controlling for random sources of variability. The key results from the model are the slope of the line between SUBJ and OBJ conditions in each experiment. The slope coefficient for Experiment

4 was -0.213, for Experiment 2 was -0.0242 and for Experiment 5 was 0.052. These represent the difference in acceptability Z score within each experiment between the SUBJ and OBJ conditions. The steady increase in acceptability as the experiments increased filler complexity and filler-gap head Pointwise Mutual Information confirms that these manipulations mitigated the acceptability decrease introduced by extractions from a complex subject. For the details of the model refer to Appendix B.

## 5 Complex Subjects and Verb Class

## 5.1 Overview

These experiments explore another potential factor that has been proposed to influence the acceptability of extractions from complex subjects in the work of Kluender (Kluender 2005). Kluender suggests that the type of verb in the complex subject may influence acceptability. Kluender observes that extractions from the beneficiary position of gerundive verbs like *bake* in complex subjects give rise to improved acceptability compared to extractions from other thematic positions. He hypothesizes that this could either be due to a preference for extraction from thematic roles lower on the thematic hierarchy or to a simple preference for a longer distance between the filler and the gap in the complex subject. To test the linear distance hypothesis we will examine verbs of correspondence such as *talk*, which take a prepositional complement. Crucially, the distance between the site of extraction and the filler for these verbs lies at an intermediate linear distance between the transitive verbs used in Experiment 2 and the benefactive verbs used in Experiment 6.

## 5.2 Experiment 6: Extraction From Benefactive Complex Subjects

### 5.2.1 Design

Experiment 6 is based on the familiar framework used in Experiment 2 with the replacement of the transitive verbs with verbs drawn from the benefactive class outlined in Levin (2000). The gap was included in the position of the *beneficiary* for the SUBJ condition. The experiment is a 2x3 design with two levels manipulating the site of extraction, either the beneficiary of the complex subject or the direct object of the matrix verb (SUBJ, OBJ) and three levels manipulating the subject of the complex subject (POSS, NULL and NAME). Each subject rated the acceptability of 20 randomly ordered sentences including 12 target sentences and 8 filler sentences. The 12 target sentences contained one sentence built from each template and two sentences from each condition and were randomized across subjects.

#### 5.2.2 Material

72 stimuli were produced from 12 templates by manipulating the 6 experimental conditions. Templates consisted of a gerundive benefactive verb, a paired noun, a paired direct object, a proper name and either the matrix verb *bother* or *upset*. Direct objects were selected by choosing high frequency trigrams beginning with the benefactive verb and ending with an indefinite NP, e.g. *baking a cake*. Target sentences were constructed by selecting the gap location; either the beneficiary of the complex gerundive subject (SUBJ) or the direct object of the matrix verb (OBJ). The complex NP subject was completed by selecting between a possessive subject my (POSS), a null subject (NULL), or a distinct proper name (NAME) for the gerundive NP. Examples of all six conditions for one template follow. A full stimulus list can be found in Appendix A.

**Template:** writing, commissioner, a book, Joe, bother.

SUBJxPOSS: Which commissioner would my writing a book for bother Joe?

SUBJxNULL: Which commissioner would writing a book for bother Joe?

**SUBJxNAME:** Which commissioner would Steve writing a book for bother Joe?

**OBJxPOSS:** Which commissioner would my writing a book for Joe bother?

**OBJxNULL:** Which commissioner would writing a book for Joe bother?

**OBJxNAME:** Which commisioner would Steve writing a book for Joe bother?

#### 5.2.3 Participants

90 participants who declared themselves native speakers of English and connected from an IP address located in the United States completed the experiment in an average of 209 seconds each.

#### 5.2.4 Results

Overall participants rated target sentences (Mean = 2.83, SD = 1.79) no differently then target sentences in Experiment 2 (Mean = 2.83, SD = 1.83, p = .95). Subjects rated extractions from benefactive complex subjects (Mean = 2.73, SD = 1.76) no differently then extractions from the matrix direct object (Mean = 2.94, SD = 1.81, p = 0.07), although this difference approached significance. A more sophisticated analysis will explore this difference in more detail in section 5.4.

## 5.3 Experiment 7: Extraction From Correspondence Complex Subjects

### 5.3.1 Design

Experiment 7 again replaces the transitive verbs used in Experiment 2 with verbs of correspondence such as *cooperate* collected by Levin (Levin 1993). All verbs took a *with* prepositional phrase complement, and the gap was included as the object of the prepositional phrase in the SUBJ condition. The experiment is a 2x3 design with two levels manipulating the site of extraction, either the prepositional object of the complex subject or the direct object of the matrix verb (SUBJ, OBJ) and three levels manipulating the subject of the complex subject (POSS, NULL and NAME). Each subject rated the acceptability of 20 randomly ordered sentences including 12 target sentences and 8 filler sentences. The 12 target sentences contained one sentence built from each template and two sentences from each condition and were randomized across subjects.

#### 5.3.2 Material

72 stimuli were produced from 12 templates by manipulating the six experimental conditions. Templates consisted of a NP headed by a gerundive correspondence verb, a paired noun, a proper name and either the matrix verb *bother* or *upset*. Target sentences were constructed by selecting the gap location, either the object of the prepositional phrase in the complex subject (SUBJ) or the direct object of the matrix verb (OBJ). We then completed the complex subject by selecting between a possessive subject my (POSS), a null subject (NULL), or a distinct proper name (NAME) for the complex NP. Examples of all six conditions for one template follow. A full stimulus list can be found in Appendix A.

Template: cooperating with, operative, John, bother.

SUBJxPOSS: Which operative would my cooperating with bother John?

SUBJxNULL: Which operative would cooperating with bother John?

SUBJxNAME: Which operative would Joe cooperating with bother John?

**OBJxPOSS:** Which operative would my cooperating with John bother?

**OBJxNULL:** Which operative would cooperating with John bother?

**OBJxNAME:** Which operative would Joe cooperating with John bother?

#### 5.3.3 Participants

90 participants who declared themselves native speakers of English and connected from an IP address located in the United States completed the experiment in an average of 202 seconds each.

#### 5.3.4 Results

Participants rated target sentences (Mean = 3.39, SD = 1.74) significantly higher than target sentences from Experiment 2 (Mean = 2.83, SD = 1.83, p < .05) and Experiment 6 (Mean = 2.83, SD = 1.79, p < .05). Participants rated extractions from the prepositional object of correspondence complex subjects (Mean = 3.42, SD = 1.80) no differently than matrix direct object extractions (Mean = 3.36, SD = 1.68, p = .55).

### 5.4 Discussion

Experiments 6 and 7 explored the effects of the verb class of the head of the complex subject on the acceptability of sentences containing complex subjects and filler-gap dependencies. The experiments also demonstrated the difference in acceptability ratings between extractions from complex subjects built around these verb types and extractions from matrix direct objects. Figure 3a gives a comparison between the distribution of acceptability ratings across Experiment 2 (mi), Experiment 6 (ben) and Experiment 7 (corr) broken down by SUBJ (Blue) or OBJ (Green) conditions and Figure 3b provides a similar perspective after z-score normalizing the ratings. We see little difference between the overall rating distributions between Experiment 6 (ben) and Experiment 2 (mi), although Figure 3b illustrates the greater difference in acceptability between SUBJ and OBJ extractions in these experiments. Experiment 7 (corr) shows a significantly different distribution from all other experiments involving a complex subject extraction, with the majority of ratings higher on the acceptability scale. The complete lack of difference between SUBJ and OBJ extraction conditions in Experiment 7 is evident in Figure 3a and even more clearly in 3b.

Experiment 6 demonstrated no improvement in acceptability when extracting from the beneficiary position of the complex subject compared to the direct object position. This is contrary to the observations in (Kluender 2005) and is problematic for his linear distance hypothesis and his thematic role hypothesis. In both hypotheses, complex subject extractions should be significantly better in Experiment 6 than in Experiment 2. An alternative explanation relies on the fact that benefactive verbs have more subcategorization options than the transitive verbs used in Experiment 2 and the correspondence verbs used in Experiment 7. The uncertainty surrounding the syntactic structure following the benefactive verb could increase the processing complexity enough to remove any benefit associated with increased distance between the filler and the gap or the extraction from a thematic role lower on the thematic heirarchy. It should be noted that the Pointwise Mutual Information present between the benefactive verb and the filler was not controlled in these experiments and is likely significantly lower than in Experiment 2, which would also reduce acceptability.

Experiment 7 demonstrated a significant improvement in complex subject extraction acceptability compared to all previous experiments. On average participants rated complex subject extractions no different from matrix direct object extractions. This non-structural manipulation that removes the difference in





acceptability between complex subject gaps and matrix object gaps give further evidence that is difficult to reconcile with a categorical prohibition on extractions from complex subjects. It is equally difficult to reconcile with Kluender's hypothesis about the source of increased acceptability. While the structural factors of having a complex subject and an unbounded dependency do contribute to the generally low acceptability of these sentences, we have evidence that they are not the only factors involved and that the structural distinction between having a gap within the complex subject or outside the complex subject has been over emphasized as the source of the difference in acceptability. As mentioned above, an alternative explanation could have to do with a balance in the processing costs associated with introducing a complex subject and the uncertainty with respect to its subcategorization frame. The parser must recognize the presence of a complex subject at the benefactive, transitive or correspondence verb. This will incur significant processing costs associated generally with complex subjects. As noted in a number of self-paced reading experiments, this complexity often carries over to subsequent words. An increase in linear distance between the complex subject verb and the gap position would give the parser time to recover from recognizing the presence of a complex subject in time to process the filler-gap dependency. Working against the benefit of linear distance is the fact that increased subcategorization options generally increase processing complexity by introducing uncertainty with respect to the syntactic structure that follows the verb and the thematic roles a particular subcategorization frame requires. In the case of the benefactive NP, the subcategorization uncertainty could have offset the increased complex subject recovery time. In Experiment 7 the simple addition of a preposition could have provided enough time for the parser to recover, and the limited subcategorization uncertainty could have allowed a relatively easy filler-gap dependency formation, ultimately resulting in improved acceptability. Further exploration into the time series of processing complex subjects and filler-gap dependencies they participate in is required to determine the precise source of increased acceptability correspondence and benefactive verb classes provide. Additional verb classes should also be explored to determine if subcategorization frame frequency, and especially their frequency of appearing in complex subjects, influences acceptability and processing complexity.

We fit a Linear Mixed Effects model including experiment as a fixed effect and allowing the slope and intercept of the subject extraction SUBJ condition to vary across experiments. For Experiment 6 the slope coefficient was 0.058 while for Experiment 7 the coefficient was 0.130. These coefficients reinforce the results previously discussed. According to the model, there was a slight increase in acceptability for the SUBJ condition in Experiment 6, but this is dwarfed by the increase found in Experiment 7. For the details of the model refer to Appendix B.

## 6 Complex Subjects and Discourse Referents

## 6.1 Overview

We have observed the effects of informativity and verb class on acceptability. We now turn to discourse referents, a final factor known to influence processing complexity and present as a dormant condition in the previous experiments. A large body of research recognizes that introducing a new discourse referent into the discourse increases processing load (Warren & Gibson 2002). Researchers also recognize a giveness hierarchy where certain discourse referents are more accessible than others (Gundel et al. 1993, Ariel 1999). Increased accessibility is associated with decreased processing costs. At the top of most giveness hierarchies are the discourse participants themselves, while newly mentioned proper names rate low on the hierarchy. We relate these findings to our current experiments by manipulating the subject NP within the complex subject. Any manipulation which reduces the overall processing load of extractions from complex subjects should result in an increase in acceptability. The high accessibility condition (POSS) makes use of the possessive pronoun My, which should have low processing complexity due to its status as a first person pronoun while the low accessibility condition (NAME) makes use of a proper name such as Joe and should result in increased processing complexity due to its new status in the discourse. We also include the null subject case (NULL) to test the observations in Kluender (2005), which hold that the null subject condition should lead to increased acceptability over the overt subject conditions. For this analysis we collapse results across Experiments 2, 4, 5, 6 and 7. Refer to the Materials sections of the previous experiments for example sentences in each condition.

## 6.2 Results

Overall participants rated questions with possessive pronoun subjects (POSS) in the complex subject (Mean = 3.06, SD = 1.86) no different from the null subject in the complex subject (NULL) condition (Mean = 2.99, SD = 1.86, p = 0.31), but significantly higher than the proper name in the complex subject (NAME) condition (Mean = 2.84, SD = 1.79, p < .05). This is confirmed by a Linear Mixed Effects model which includes the subject type in the complex subject as a fixed effect. Using the (POSS) condition as the basis the coefficient for the presence of the (NULL) condition was -0.005 while the coefficient for the (NAME) condition was -0.114. For the complete results of the Linear Mixed Effects model see Appendix B.

### 6.3 Discussion

This section explored the effect of discourse reference accessibility on the acceptability of complex subject extractions. We observed that inclusion of a possessive pronoun or a null subject as the subject of the complex subject improved acceptability judgements compared to the use of a proper name. Figure 4a gives a comparison between the distribution of acceptability ratings across subject types in the complex subject (POSS, NULL, NAME) broken down by SUBJ(Blue) or OBJ(Green) conditions. Figure 4b provides a similar perspective after z-score normalization of the ratings. We can observe the slight increase in mass to the right of the graph in the POSS and NULL conditions in Figure 4b. We also observe the smallest difference between SUBJ and OBJ conditions in the NAME experimental condition.

These results are consistent with previous results on the processing complexity of discourse referents based on the their accessibility. Contrary to the claims in Kluender (2004), we observed little difference between the use of a null subject or a possessive pronoun. This might arise because sentences containing a null subject in the complex subject are semantically underspecified and can receive multiple interpretations. Participants likely interpret these sentences using the most accessible discourse referent that makes semantic sense, and we should expect low processing costs and improved acceptability as a result.

It is somewhat surprising to find that including a proper name as the subject in a complex subject reduces the acceptability difference between extractions from the complex subject and extractions from the matrix object. Since proper names should incur the highest processing cost, we should see reduced acceptability when the parser tries to also perform the complex task of resolving a filler-gap dependency inside the complex subject. To the contrary we actually see an increase in acceptability for the SUBJ condition as confirmed by the positive interaction between the SUBJ and NAME conditions in our Linear Mixed Effects model. We currently have no explanation for this result although we should note that the interaction coefficient is small and may be an artifact of collapsing across all experiments.

## 7 Conclusion

This research demonstrated two well-known processing factors that influence the acceptability of Complex Subject Island violations. As processing complexity is reduced, either by increasing filler informativity or discourse referent accessibility, the acceptability of the island violation increases. As processing complexity is increased, the result on acceptability is the opposite. We identified an additional non-structural factor, the verb class within the Complex Subject, that impacts the acceptability of Complex Subject Island violations. We have demonstrated that declarative sentences with complex subjects are deemed slightly less acceptable than simple wh-questions without island constraint violations, confirming the findings that complex subjects are dispreferred in both comprehension and production. We further demonstrated that the co-occurrence of two syntactic structures that are difficult to process, Complex Subjects and filler-gap dependencies significantly reduces accptability regardless of whether the gap is located within the syntactic island or not. We confirm that when the gap is located within the complex, subject acceptability is further reduced and demonstrate how to systematically mitigate this difference by manipulating Mu-





tual Information between the filler and the gap, the informativity of the filler, the complex subject verb class and the accessibility of interveining discourse referents.

The graded nature of the acceptability results suggests that these phenomena cannot be accounted for as a categorical prohibition on syntactic configurations in the competence grammar. The strong interaction with known processing factors suggests that these findings are amenable to a processing account. Such an account must predict the individual processing difficulty of rare syntactic structures such as Complex Subjects and complex syntactic structures like unbounded filler-gap dependencies. The account must also take into account the interaction of the various sources of processing complexity highlighted in this research.

## 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.
- [2] Almor, A. (2004). A computational investigation of reference in production and comprehension. In J. Trueswell & M. Tanenhaus (Eds.), Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions (pp. 285–301). Cambridge, MA: MIT Press.
- [3] Ariel, M. (1999). Cognitive Universals and Linguistic Conventions: The Case of Resumptive Pronouns. Studies in Language 23:2, 217-269.
- [4] Arnon, I., Hofmeister, P., Jaeger, T. F., Sag, I., & Snider, N. (2005). Rethinking superiority effects: a processing model. Poster presented at the CUNY Sentence Processing Conference.
- [5] Bates, Douglas M. (2005). Fitting Linear Mixed Models in R. R News, 5(1), 27-30.
- [6] Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. Journal of Memory and Language, 52, 144-169.
- [7] Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), A Festschrift for Morris Halle (p. 232-86). New York: Holt, Reinhart & Winston.
- [8] Chomsky, N. (1977). On wh-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), Formal Syntax. New York: Academic Press.

- [9] Chomsky, N. (1981). Lectures on Government and Binding. Dordrecht: Foris.
- [10] Chomsky, N. (1986). Barriers. Cambridge: MIT Press.
- [11] Cinque, G. (1990). Types of A'-dependencies. Cambridge: MIT Press.
- [12] Clifton, C., & Frazier, L. (1989). Comprehending sentences with longdistance dependencies. In M. Tanenhaus & G. Carlson (Eds.), Linguistic Structure in Language Processing (pp. 273–317). Dordrecht: Kluwer Academic Publishers.
- [13] Cowart, W. (1997). Experimental Syntax: Applying Objective Methods in Sentence Judgments. Thousand Oaks, CA: Sage Publications.
- [14] Deane, P. (1991). Limits to attention: a cognitive theory of island phenomena. Cognitive Linguistics, 2 (1), 1-63.
- [15] De Vincenzi, M. (1996). Syntactic analysis in sentence comprehension: Effects of dependency types and grammatical constraints. Journal of Psycholinguistic Research, 25 (1), 117–133.
- [16] Erteschik-Shir, N. (1973). On the Nature of island constraints. Unpublished doctoral dissertation, MIT.
- [17] Erteschik-Shir, N. (2004). What's what? In G. Fanselow, C. Fery, M. Schlesewsky, & R. Vogel (Eds.), Gradience in Grammar. Oxford: Oxford University Press.
- [18] Erteschik-Shir, N., & Lappin, S. (1979). Dominance and the functional explanation of island constraints. Theoretical Linguistics, 6 (1), 43–84.
- [19] Fanselow, G., & Frisch, S. (2004). Effects of processing difficulty on judgments of acceptability. In G. Fanselow, C. Fery, M. Schlesewsky, & R. Vogel (Eds.), Gradience in Grammar. Oxford: Oxford University Press.
- [20] Fiebach, C., Schlesewsky, M., & Friederici, A. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. Journal of Psycholinguistic Research, 30 (3), 321–338.
- [21] Gibson, E. (1991). A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.
- [22] Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. Cognition, 68 (1), 1-76.
- [23] Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), Image, Language, Brain (p. 95-126). Cambridge: MIT Press.

- [24] Gibson, E., & Pearlmutter, N. (1998). constraints on sentence comprehension. Trends in Cognitive Science, 2, 262–268.
- [25] Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. Language and Cognitive Processes, 14 (3), 225–248.
- [26] Gordon, P., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. Journal of Memory and Language, 51, 97-114.
- [27] Gordon, P., Hendrick, R., & Levine, W. (2002). Memory-load interference in syntactic processing. Psychological Science, 13, 425-430.
- [28] Graff, D., Cieri, C. (2003). English Gigaword. Linguistic Data Consortium, Philadelphia.
- [29] Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. Cognitive Science, 29, 261–290.
- [30] Gundel, J., Hedberg, N., Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. Language. 69:2, 274-307.
- [31] Haegeman, L. (1994). Introduction to Government and Binding Theory. Oxford: Blackwell Publishing.
- [32] Hawkins, J. (1999). Processing complexity and filler-gap dependencies across grammars. Language, 75, 244-285.
- [33] Hickok, G., & Avrutin, S. (1995). Representation, referentiality, and processing in agrammatic comprehension: two case studies. Brain and Language, 50 (1), 10–26.
- [34] Hofmeister, P. (2007). Retrievability and gradience in filler-gap dependencies. In Proceedings of the 43 Annual Meeting of the Chicago Linguistic Society. Chicago: University of Chicago.
- [35] Hofmeister, Philip. (2008). Representational Complexity and Memory Retrieval in Language Comprehension. Stanford University doctoral dissertation.
- [36] Hofmeister, P., Jaeger, T., Sag, I., Arnon, I., & Snider, N. (2007). Locality and accessibility in wh-questions. In S. Featherston & W. Sternefeld (Eds.), Roots: Linguistics in Search of its Evidential Base. Berlin: Mouton de Gruyter.
- [37] Hofmeister, P., Arnon, I., Jaeger, T., Sag, I., Snider, N. (2010). The source ambiguity problem: distinguishing the effects of grammar and processing on acceptability judgements. Unpublished Manuscript. Stanford University.
- [38] Hofmeister, P. Sag I. (2010). Cognitive constraints on syntactic islands. In Press. Stanford University.

- [39] Arnon, I., Snider, S., Hofmeister, P., Jaeger T., & Sag, A. Cross-linguistic Variation in a Processing Account: The Case of Multiple Wh-questions. Unpublished Manuscript. Stanford University.
- [40] Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), Probablistic Linguistics. Boston: MIT Press.
- [41] Keller, F. (2004). The entropy rate principle as a predictor of processing effort: An evaluation against eye-tracking data. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Barcelona, pp. 317-324.
- [42] Kluender, R. (1991). Cognitive constraints on Variables in Syntax. Unpublished doctoral dissertation, University of California, San Diego.
- [43] Kluender, R. (1992). Deriving islands constraints from principles of predication. In H. Goodluck & M. Rochemont (Eds.), Island Constraints: Theory, Acquisition and Processing (p. 223-258). Dordrecht: Kluwer.
- [44] Kluender, R. (1998). On the distinction between strong and weak islands: a processing perspective. In P. Culicover & L. McNally (Eds.), Syntax and Semantics 29: The Limits of Syntax (p. 241-279). San Diego, CA: Academic Press.
- [45] Kluender, R. (2005). Are subject islands subject to a processing account? Proceedings of the 23rd West Coast Conference on Formal Linguistics, 475–499.
- [46] Kluender, R., & Kutas, M. (1993a). Bridging the gap: evidence from erps on the processing of unbounded dependencies. Journal of Cognitive Neuroscience, 5 (2), 196-214.
- [47] Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. Language and Cognitive Processes, 8 (4).
- [48] Konieczny, L. (2000). Locality and parsing complexity. Journal of Psycholinguistic Research, 29, 627-645.
- [49] Kravtchenko, E., Polinsky M., Xiang M (2009). Are All Subject Islands Created Equal? CUNY Poster Presentation.
- [50] Lee, M. (2004). Another Look at the Role of Empty Categories in Sentence Processing (and Grammar). Journal of Psycholinguistic Research, 33(1), 51-73.
- [51] Levin, B. (1993). English Verb Classes and Alternations. Chicago: University of Chicago Press.
- [52] Levy, R. (2005). Probalistic Models of Word Order and Syntactic Discontinuity. Unpublished doctoral dissertation, Stanford University.

- [53] Levy, R. (2007). Expectation-based syntactic comprehension. Unpublished Manuscript. University of California, San Diego.
- [54] Maling, J. (1978). An asymmetry with respect to wh-islands. Linguistic Inquiry, 9, 75-89.
- [55] McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. Journal of Experimental Psychology. Learning, Memory, and Cognition, 24 (2), 432–460.
- [56] Munro, R., Bethard, S., Kuperman, V., Melnick, R., Potts, C., Schnoebelen, T., Tily, H. (2010). Crowdsourcing and Language Studies: The new Generation of Linguistic Data. NAACL 2010 Workshop on Creating Speech and Language Data with Mechanical Turk.
- [57] Nunes, J., & Uriagereka, J. (2000). Cyclicity and Extraction Domains. Syntax, 3:1, 20-43.
- [58] Oshima, D. (2007). On factive islands: pragmatic anomaly vs. pragmatic infelicity. In T. Washio, K. Satoh, H. Terada, & A. Inokuchi (Eds.), New Frontiers in Artificial Intelligence: Joint JSAI 2006 Workshop Post-Proceedings (pp. 147–161). Dordrecht: Springer-Verlag.
- [59] Pesetsky, D. (2000). Phrasal Movement and its Kin. Cambridge: MIT Press.
- [60] Phillips, C. , Kazanina, N., Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Research, 22, 407– 428.
- [61] Phillips, C. (2006). The Real-Times Status of island constraints Rizzi, L. (1982). Violations of the wh-island constraint and the Subjacency Condition. Issues in Italian Syntax, 49–76.
- [62] Pickering, M. (1994). Processing Local and Unbounded Dependencies: A Unified Account. Journal of Psycholinguistic Research, 23:4, 323-352.
- [63] Pickering, M., Traxler, M. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. Language and Cognitive Processes 18:4, 469-503.
- [64] Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. Cognition, 61, 127-159.
- [65] Rizzi, L. (1990). Relativized Minimality. Cambridge: MIT Press.
- [66] Rizzi, L. (2000). Reconstruction, Weak island Sensitivity, and Agreement. (Ms. Universit'a di Siena)
- [67] Ross, J. R. (1967). constraints on Variables in Syntax. PhD Thesis. MIT.

- [68] Saah, K., & Goodluck, H. (1995). island effects in parsing and grammar: evidence from Akan. Linguistic Review, 12 (4), 381–409.
- [69] Sabel, J. (2002a). A minimalist analysis of syntactic islands. The Linguistic Review, 19, 271–315.
- [70] Sag, I. A., Hofmeister, P., & Snider, N. (2007). Processing complexity in Subjacency violations: the complex noun phrase constraint. In Proceedings of the 43 Annual Meeting of the Chicago Linguistic Society. Chicago: University of Chicago.
- [71] Schütze, C. (1996). The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology. Chicago: University of Chicago Press.
- [72] Sprouse, J. (2008). The Differential Sensitivity of Acceptability Judgements to Processing Effects. Squib.
- [73] Sussman, R., Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. Language and Cognitive Processes 18 (2), 143-163.
- [74] Szabolcsi, A. (1992). Weak islands, individuals, and scope. In Proceedings of SALT II. Ohio State University, Columbus.
- [75] Szabolcsi, A. (2006). Strong vs. weak islands. The Blackwell Companion to Syntax, 4, 479–531.
- [76] Szabolcsi, A., & Zwarts, F. (1993). Weak islands and an algebraic semantics of scope taking. Natural Language Semantics, 1, 235-284.
- [77] Traxler, M., & Pickering, M. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. Journal of Memory and Language, 35 (3), 454–475.
- [78] Truswell, R. (2007). Locality of *Wh*-movement and the Individuation of Events. Unpublished doctoral dissertation, University College London.
- [79] Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33 (2), 407-430.
- [80] Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. Language, 82 (3).
- [81] Wagers, M. (2008). The Structure of Memory Meets Memory for Structure in Linguistic Cognition. Unpublished Doctoral Dissertation. University of Maryland, College Park.

- [82] Wagers, M., Phillips, C. (2009). Multiple dependencies and the role of grammar in real-time comprehension. Journal of Linguistics, 45:2, 395-433.
- [83] Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. Cognition, 85, 79-112.

## 8 Appendix A

## 8.1 Experiment 1 Stimuli

My deceiving a customer would bother Sarah. My disciplining a student would upset Chris. My arresting a perpetrator would bother Susan. My enlisting a partner would upset Steve. My appointing a commissioner would bother Joe. My intimidating a scientist would upset Greg. My teaching a pupil would bother Janet. My charging a juvenile would upset Laura. My hiring a planner would bother Ben. My seducing a girl would upset Kate. My hiring an operative would bother John. My mistreating a child would upset Megan. My deceiving Sarah would bother a customer. My disciplining Chris would upset a student. My arresting Susan would bother a perpetrator. My enlisting Steve would upset a partner. My appointing Joe would bother a commissioner. My intimidating Greg would upset a scientist. My teaching Janet would bother a pupil. My charging Laura would upset a juvenile. My hiring Ben would bother a planner. My seducing Kate would upset a girl. My hiring John would bother an operative. My mistreating Megan would upset a child. Deceiving a customer would bother Sarah. Disciplining a student would upset Chris. Arresting a perpetrator would bother Susan. Enlisting a partner would upset Steve. Appointing a commissioner would bother Joe.

Intimidating a scientist would upset Greg. Teaching a pupil would bother Janet. Charging a juvenile would upset Laura. Hiring a planner would bother Ben. Seducing a girl would upset Kate. Hiring an operative would bother John. Mistreating a child would upset Megan.

Deceiving Sarah would bother a customer. Disciplining Chris would upset a student. Arresting Susan would bother a perpetrator. Enlisting Steve would upset a partner. Appointing Joe would bother a commissioner. Intimidating Greg would upset a scientist. Teaching Janet would bother a pupil. Charging Laura would upset a juvenile. Hiring Ben would bother a planner. Seducing Kate would upset a girl. Hiring John would bother an operative. Mistreating Megan would upset a child.

Megan deceiving a customer would bother Sarah. John disciplining a student would upset Chris. Chris arresting a perpetrator would bother Susan. Susan enlisting a partner would upset Steve. Steve appointing a commissioner would bother Joe. Kate intimidating a scientist would upset Greg. Laura teaching a pupil would bother Janet. Ben charging a juvenile would upset Laura. Janet hiring a planner would bother Ben. Greg seducing a girl would upset Kate. Joe hiring an operative would bother John. Sarah mistreating a child would upset Megan.

Megan deceiving Sarah would bother a customer. John disciplining Chris would upset a student. Chris arresting Susan would bother a perpetrator. Susan enlisting Steve would upset a partner. Steve appointing Joe would bother a commissioner. Kate intimidating Greg would upset a scientist. Laura teaching Janet would bother a pupil. Ben charging Laura would upset a juvenile. Janet hiring Ben would bother a planner. Greg seducing Kate would upset a girl. Joe hiring John would bother an operative. Sarah mistreating Megan would upset a child.

## 8.2 Experiment 2 Stimuli:

Which customer would my deceiving bother Sarah? Which student would my disciplining upset Chris? Which perpetrator would my arresting bother Susan? Which partner would my enlisting upset Steve? Which commissioner would my appointing bother Joe? Which scientist would my intimidating upset Greg? Which pupil would my teaching bother Janet? Which juvenile would my charging upset Laura? Which planner would my hiring bother Ben? Which girl would my seducing upset Kate? Which operative would my hiring bother John? Which child would my mistreating upset Megan?

Which customer would my deceiving Sarah bother? Which student would my disciplining Chris upset? Which perpetrator would my arresting Susan bother? Which partner would my enlisting Steve upset? Which commissioner would my appointing Joe bother? Which scientist would my intimidating Greg upset? Which pupil would my teaching Janet bother? Which juvenile would my charging Laura upset? Which planner would my hiring Ben bother? Which girl would my seducing Kate upset? Which operative would my hiring John bother? Which child would my mistreating Megan upset?

Which customer would deceiving bother Sarah? Which student would disciplining upset Chris? Which perpetrator would arresting bother Susan? Which partner would enlisting upset Steve? Which commissioner would appointing bother Joe? Which scientist would intimidating upset Greg? Which pupil would teaching bother Janet? Which juvenile would charging upset Laura? Which planner would hiring bother Ben? Which girl would seducing upset Kate? Which operative would hiring bother John? Which child would mistreating upset Megan?

Which customer would deceiving Sarah bother? Which student would disciplining Chris upset? Which perpetrator would arresting Susan bother? Which partner would enlisting Steve upset? Which commissioner would appointing Joe bother? Which scientist would intimidating Greg upset? Which pupil would teaching Janet bother? Which juvenile would charging Laura upset? Which planner would hiring Ben bother? Which girl would seducing Kate upset? Which operative would hiring John bother? Which child would mistreating Megan upset?

Which customer would Megan deceiving bother Sarah? Which student would John disciplining upset Chris? Which perpetrator would Chris arresting bother Susan? Which partner would Susan enlisting upset Steve? Which commissioner would Steve appointing bother Joe? Which scientist would Kate intimidating upset Greg? Which pupil would Laura teaching bother Janet? Which juvenile would Ben charging upset Laura? Which planner would Janet hiring bother Ben? Which girl would Greg seducing upset Kate? Which operative would Joe hiring bother John? Which child would Sarah mistreating upset Megan?

Which customer would Megan deceiving Sarah bother? Which student would John disciplining Chris upset? Which perpetrator would Chris arresting Susan bother? Which partner would Susan enlisting Steve upset? Which commissioner would Steve appointing Joe bother? Which scientist would Kate intimidating Greg upset? Which pupil would Laura teaching Janet bother? Which juvenile would Ben charging Laura upset? Which planner would Janet hiring Ben bother? Which girl would Greg seducing Kate upset? Which operative would Joe hiring John bother? Which child would Sarah mistreating Megan upset?

### 8.3 Experiment 3 Stimuli

Which customer would Sarah regret my deceiving? Which student would Chris dislike my disciplining? Which perpetrator would Susan regret my arresting? Which partner would Steve dislike my enlisting? Which commissioner would Joe regret my appointing? Which scientist would Greg dislike my intimidating? Which pupil would Janet regret my teaching? Which juvenile would Laura dislike my charging? Which planner would Ben regret my hiring? Which girl would Kate dislike my seducing? Which operative would John regret my hiring? Which child would Megan dislike my mistreating?

Which customer would regret my deceiving Sarah?

Which student would dislike my disciplining Chris? Which perpetrator would regret my arresting Susan? Which partner would dislike my enlisting Steve? Which commissioner would regret my appointing Joe? Which scientist would dislike my intimidating Greg? Which pupil would regret my teaching Janet? Which juvenile would dislike my charging Laura? Which planner would regret my hiring Ben? Which girl would dislike my seducing Kate? Which operative would regret my hiring John? Which child would dislike my mistreating Megan?

Which customer would Sarah regret deceiving? Which student would Chris dislike disciplining? Which perpetrator would Susan regret arresting? Which partner would Steve dislike enlisting? Which commissioner would Joe regret appointing? Which scientist would Greg dislike intimidating? Which pupil would Janet regret teaching? Which juvenile would Laura dislike charging? Which planner would Ben regret hiring? Which girl would Kate dislike seducing? Which operative would John regret hiring? Which child would Megan dislike mistreating?

Which customer would regret deceiving Sarah? Which student would dislike disciplining Chris? Which perpetrator would regret arresting Susan? Which partner would dislike enlisting Steve? Which commissioner would regret appointing Joe? Which scientist would dislike intimidating Greg? Which pupil would regret teaching Janet? Which juvenile would dislike charging Laura? Which planner would regret hiring Ben? Which girl would dislike seducing Kate? Which operative would regret hiring John? Which child would dislike mistreating Megan? Which customer would Sarah regret Megan deceiving?

Which student would Chris dislike John disciplining? Which perpetrator would Susan regret Chris arresting? Which partner would Steve dislike Susan enlisting? Which commissioner would Joe regret Steve appointing? Which scientist would Greg dislike Kate intimidating? Which pupil would Janet regret Laura teaching? Which juvenile would Laura dislike Ben charging? Which planner would Ben regret Janet hiring? Which girl would Kate dislike Greg seducing? Which operative would John regret Joe hiring? Which child would Megan dislike Sarah mistreating?

Which customer would regret Megan deceiving Sarah? Which student would dislike John disciplining Chris? Which perpetrator would regret Chris arresting Susan? Which partner would dislike Susan enlisting Steve? Which commissioner would regret Steve appointing Joe? Which scientist would dislike Kate intimidating Greg? Which pupil would regret Laura teaching Janet? Which juvenile would dislike Ben charging Laura? Which planner would regret Janet hiring Ben? Which girl would dislike Greg seducing Kate? Which operative would regret Joe hiring John? Which child would dislike Sarah mistreating Megan?

### 8.4 Experiment 4 Stimuli

Who would my deceiving bother Sarah? Who would my disciplining upset Chris? Who would my arresting bother Susan? Who would my enlisting upset Steve? Who would my appointing bother Joe? Who would my intimidating upset Greg? Who would my teaching bother Janet? Who would my teaching upset Laura? Who would my charging upset Laura? Who would my hiring bother Ben? Who would my seducing upset Kate? Who would my hiring bother John? Who would my mistreating upset Megan?

Who would my deceiving Sarah bother? Who would my disciplining Chris upset? Who would my arresting Susan bother? Who would my enlisting Steve upset? Who would my appointing Joe bother? Who would my intimidating Greg upset? Who would my teaching Janet bother? Who would my teaching Laura upset? Who would my charging Laura upset? Who would my hiring Ben bother? Who would my seducing Kate upset? Who would my hiring John bother? Who would my mistreating Megan upset? Who would deceiving bother Sarah? Who would disciplining upset Chris? Who would arresting bother Susan? Who would enlisting upset Steve? Who would appointing bother Joe? Who would intimidating upset Greg? Who would teaching bother Janet? Who would charging upset Laura? Who would hiring bother Ben? Who would seducing upset Kate? Who would hiring bother John? Who would mistreating upset Megan?

Who would deceiving Sarah bother? Who would disciplining Chris upset? Who would arresting Susan bother? Who would enlisting Steve upset? Who would appointing Joe bother? Who would intimidating Greg upset? Who would teaching Janet bother? Who would charging Laura upset? Who would hiring Ben bother? Who would seducing Kate upset? Who would hiring John bother? Who would hiring John bother?

Who would Megan deceiving bother Sarah? Who would John disciplining upset Chris? Who would Chris arresting bother Susan? Who would Susan enlisting upset Steve? Who would Steve appointing bother Joe? Who would Kate intimidating upset Greg? Who would Laura teaching bother Janet? Who would Ben charging upset Laura? Who would Ben charging upset Laura? Who would Greg seducing upset Kate? Who would Joe hiring bother John? Who would Sarah mistreating upset Megan?

Who would Megan deceiving Sarah bother? Who would John disciplining Chris upset? Who would Chris arresting Susan bother? Who would Susan enlisting Steve upset? Who would Steve appointing Joe bother? Who would Kate intimidating Greg upset? Who would Laura teaching Janet bother? Who would Ben charging Laura upset? Who would Janet hiring Ben bother? Who would Greg seducing Kate upset? Who would Joe hiring John bother? Who would Sarah mistreating Megan upset?

## 8.5 Experiment 5 Stimuli

Which customer with a coupon would my deceiving bother Sarah? Which student from Prague would my disciplining upset Chris? Which perpetrator with a motive would my arresting bother Susan?

Which partner from Turkey would my enlisting upset Steve?

Which commissioner with a question would my appointing bother Joe?

Which scientist from New Jersey would my intimidating upset Greg?

Which pupil with Aspergers would my teaching bother Janet? Which juvenile from Kansas would my charging upset Laura? Which planner with experience would my hiring bother Ben? Which girl from Monaco would my seducing upset Kate? Which operative with an agenda would my hiring bother John? Which child from Brazil would my mistreating upset Megan?

Which customer with a coupon would my deceiving Sarah bother? Which student from Prague would my disciplining Chris upset? Which perpetrator with a motive would my arresting Susan bother? Which partner from Turkey would my enlisting Steve upset? Which commissioner with a question would my appointing Joe

bother?

Which scientist from New Jersey would my intimidating Greg upset?

Which pupil with Aspergers would my teaching Janet bother? Which juvenile from Kansas would my charging Laura upset? Which planner with experience would my hiring Ben bother? Which girl from Monaco would my seducing Kate upset? Which operative with an agenda would my hiring John bother? Which child from Brazil would my mistreating Megan upset?

Which customer with a coupon would deceiving bother Sarah? Which student from Prague would disciplining upset Chris? Which perpetrator with a motive would arresting bother Susan? Which partner from Turkey would enlisting upset Steve?

Which commissioner with a question would appointing bother Joe?

Which scientist from New Jersey would intimidating upset Greg?

Which pupil with Aspergers would teaching bother Janet? Which juvenile from Kansas would charging upset Laura? Which planner with experience would hiring bother Ben? Which girl from Monaco would seducing upset Kate? Which operative with an agenda would hiring bother John? Which child from Brazil would mistreating upset Megan?

Which customer with a coupon would deceiving Sarah bother? Which student from Prague would disciplining Chris upset? Which perpetrator with a motive would arresting Susan bother? Which partner from Turkey would enlisting Steve upset? Which commissioner with a question would appointing Joe bother? Which scientist from New Jersey would intimidating Greg upset? Which pupil with Aspergers would teaching Janet bother? Which juvenile from Kansas would charging Laura upset? Which planner with experience would hiring Ben bother? Which girl from Monaco would seducing Kate upset? Which operative with an agenda would hiring John bother? Which child from Brazil would mistreating Megan upset?

Which customer with a coupon would Megan deceiving bother Sarah?

Which student from Prague would John disciplining upset Chris? Which perpetrator with a motive would Chris arresting bother Susan?

Which partner from Turkey would Susan enlisting upset Steve? Which commissioner with a question would Steve appointing bother Joe?

Which scientist from New Jersey would Kate intimidating upset Greg?

Which pupil with Aspergers would Laura teaching bother Janet? Which juvenile from Kansas would Ben charging upset Laura? Which planner with experience would Janet hiring bother Ben?

Which girl from Monaco would Greg seducing upset Kate?

Which operative with an agenda would Joe hiring bother John? Which child from Brazil would Sarah mistreating upset Megan?

Which customer with a coupon would Megan deceiving Sarah bother?

Which student from Prague would John disciplining Chris upset? Which perpetrator with a motive would Chris arresting Susan bother?

Which partner from Turkey would Susan enlisting Steve upset? Which commissioner with a question would Steve appointing Joe

bother?

Which scientist from New Jersey would Kate intimidating Greg upset?

Which pupil with Aspergers would Laura teaching Janet bother? Which juvenile from Kansas would Ben charging Laura upset? Which planner with experience would Janet hiring Ben bother? Which girl from Monaco would Greg seducing Kate upset? Which operative with an agenda would Joe hiring John bother? Which child from Brazil would Sarah mistreating Megan upset?

## 8.6 Experiment 6 Stimuli

Which customer would my brewing coffee for bother Sarah? Which student would my baking a cake for upset Chris? Which perpetrator would my cooking dinner for bother Susan? Which partner would my performing a play for upset Steve? Which commissioner would my writing a book for bother Joe? Which scientist would my compiling a report for upset Greg? Which pupil would my spinning a story for bother Janet? Which juvenile would my pouring a drink for upset Laura? Which planner would my developing an agenda for bother Ben? Which girl would my painting a picture for upset Kate? Which operative would my arranging a job for bother John? Which child would my frying an egg for upset Megan?

Which customer would my brewing coffee for Sarah bother? Which student would my baking a cake for Chris upset? Which perpetrator would my cooking dinner for Susan bother? Which partner would my performing a play for Steve upset? Which commissioner would my writing a book for Joe bother? Which scientist would my compiling a report for Greg upset? Which pupil would my spinning a story for Janet bother? Which juvenile would my pouring a drink for Laura upset? Which planner would my developing an agenda for Ben bother? Which girl would my painting a picture for Kate upset? Which operative would my arranging a job for John bother? Which child would my frying an egg for Megan upset?

Which customer would brewing coffee for bother Sarah? Which student would baking a cake for upset Chris? Which perpetrator would cooking dinner for bother Susan? Which partner would performing a play for upset Steve? Which commissioner would writing a book for bother Joe? Which scientist would compiling a report for upset Greg? Which pupil would spinning a story for bother Janet? Which juvenile would pouring a drink for upset Laura? Which planner would developing an agenda for bother Ben? Which girl would painting a picture for upset Kate? Which operative would arranging a job for bother John? Which child would frying an egg for upset Megan?

Which customer would brewing coffee for Sarah bother? Which student would baking a cake for Chris upset? Which perpetrator would cooking dinner for Susan bother? Which partner would performing a play for Steve upset? Which commissioner would writing a book for Joe bother? Which scientist would compiling a report for Greg upset? Which pupil would spinning a story for Janet bother? Which juvenile would pouring a drink for Laura upset? Which planner would developing an agenda for Ben bother? Which girl would painting a picture for Kate upset? Which operative would arranging a job for John bother? Which child would frying an egg for Megan upset?

Which customer would Megan brewing coffee for bother Sarah? Which student would John baking a cake for upset Chris? Which perpetrator would Chris cooking dinner for bother Susan? Which partner would Susan performing a play for upset Steve? Which commissioner would Steve writing a book for bother Joe? Which scientist would Kate compiling a report for upset Greg? Which pupil would Laura spinning a story for bother Janet? Which juvenile would Ben pouring a drink for upset Laura? Which planner would Janet developing an agenda for bother Ben?

Which girl would Greg painting a picture for upset Kate? Which operative would Joe arranging a job for bother John? Which child would Sarah frying an egg for upset Megan?

Which customer would Megan brewing coffee for Sarah bother? Which student would John baking a cake for Chris upset? Which perpetrator would Chris cooking dinner for Susan bother? Which partner would Susan performing a play for Steve upset? Which commissioner would Steve writing a book for Joe bother? Which scientist would Kate compiling a report for Greg upset? Which pupil would Laura spinning a story for Janet bother? Which juvenile would Ben pouring a drink for Laura upset? Which planner would Janet developing an agenda for Ben bother? Which girl would Greg painting a picture for Kate upset? Which operative would Joe arranging a job for John bother? Which child would Sarah frying an egg for Megan upset?

## 8.7 Experiment 7 Stimuli

Which customer would my bargaining with bother Sarah? Which student would my agreeing with upset Chris? Which perpetrator would my negotiating with bother Susan? Which partner would my arguing with upset Steve? Which commissioner would my disagreeing with bother Joe? Which scientist would my collaborating with upset Greg? Which pupil would my competing with bother Janet? Which juvenile would my bickering with upset Laura? Which planner would my differing with bother Ben? Which girl would my concurring with upset Kate? Which operative would my cooperating with bother John? Which child would my joking with upset Megan?

Which customer would my bargaining with Sarah bother? Which student would my agreeing with Chris upset? Which perpetrator would my negotiating with Susan bother? Which partner would my arguing with Steve upset? Which commissioner would my disagreeing with Joe bother? Which scientist would my collaborating with Greg upset? Which pupil would my competing with Janet bother? Which juvenile would my bickering with Laura upset? Which planner would my differing with Ben bother? Which girl would my concurring with Kate upset? Which operative would my cooperating with John bother? Which child would my joking with Megan upset?

Which customer would bargaining with bother Sarah? Which student would agreeing with upset Chris? Which perpetrator would negotiating with bother Susan? Which partner would arguing with upset Steve? Which commissioner would disagreeing with bother Joe? Which scientist would collaborating with upset Greg? Which pupil would competing with bother Janet? Which juvenile would bickering with upset Laura? Which planner would differing with bother Ben? Which girl would concurring with upset Kate? Which operative would cooperating with bother John? Which child would joking with upset Megan?

Which customer would bargaining with Sarah bother? Which student would agreeing with Chris upset? Which perpetrator would negotiating with Susan bother? Which partner would arguing with Steve upset? Which commissioner would disagreeing with Joe bother? Which scientist would collaborating with Greg upset? Which pupil would competing with Janet bother? Which juvenile would bickering with Laura upset? Which planner would differing with Ben bother? Which girl would concurring with Kate upset? Which operative would cooperating with John bother? Which child would joking with Megan upset?

Which customer would Megan bargaining with bother Sarah? Which student would John agreeing with upset Chris? Which perpetrator would Chris negotiating with bother Susan? Which partner would Susan arguing with upset Steve? Which commissioner would Steve disagreeing with bother Joe? Which scientist would Kate collaborating with upset Greg? Which pupil would Laura competing with bother Janet? Which juvenile would Ben bickering with upset Laura? Which planner would Janet differing with bother Ben? Which girl would Greg concurring with upset Kate? Which operative would Joe cooperating with bother John? Which child would Sarah joking with upset Megan?

Which customer would Megan bargaining with Sarah bother? Which student would John agreeing with Chris upset? Which perpetrator would Chris negotiating with Susan bother? Which partner would Susan arguing with Steve upset? Which commissioner would Steve disagreeing with Joe bother? Which scientist would Kate collaborating with Greg upset? Which pupil would Laura competing with Janet bother? Which juvenile would Ben bickering with Laura upset? Which planner would Janet differing with Ben bother? Which girl would Greg concurring with Kate upset? Which operative would Joe cooperating with John bother? Which child would Sarah joking with Megan upset?

## 9 Appendix B

We fit a linear mixed effects model using the lme4 package lmer() function in R (Bates, D M. 2004). The model with the best fit predicted the Z-score normalized sentence ratings using fixed effects of the extraction location and the accessibility of the subject of the complex subject and included an interaction between the two fixed effects. For random effects we allowed variable intercepts and slopes for the extraction location pooled by experiment, the extraction location pooled by the template sentence form. We also allowed variable intercepts

	coef.est	coef.se
(Intercept)	0.09	0.03
Extraction[T.SUBJ]	-0.09	0.07
Type[NULL]	0.00	0.02
Type[NAME]	-0.11	0.02
Extraction[T.SUBJ]:Type[NULL]	-0.07	0.03
Extraction[T.SUBJ]:Type[NAME]	0.03	0.03

Table 1: Fixed Effects

Groups	Name	Std.Dev	Corr
Subject ID	ExtractionOBJ	0.11	
	ExtractionSUBJ	0.11	-1.00
Subject ID	(Intercept)	0.00	
Template	(Intercept)	0.02	
	Extraction[T.SUBJ]	0.05	-0.30
Experiment	(Intercept)	0.06	
	Extraction[T.SUBJ]	0.14	-1.00
Residual		0.46	

 Table 2: Random Effects

for the partipant and variable slopes for the extraction location pooled by particpant allowing for variation at the participant level. The following provides a summary of the model and the relevant coefficients.

Number of observations: 4408, Groups: MD5, 347; Base, 12; Experiment, 5 AIC = 5868.2, DIC = 5765.9 deviance = 5800.0

lmer(formula=Z.Answer~Extraction+Type+Extraction:Type+(1+Extraction/Experiment)+(1+Extraction/Base)+(1/MD5)+(0+Extraction/MD5), data=targetMSD)

Template	(Intercept)	Extraction[T.SUBJ]
1	-0.012663360	0.056087311
10	0.014108728	0.003147099
11	0.003314242	-0.044837104
12	0.001749884	-0.040526769
2	0.008101561	0.005235280
3	-0.014620697	-0.024978761
4	0.015468552	-0.009679432
5	-0.003016080	0.047460707
6	0.008522197	0.046178010
7	-0.004611515	-0.042580882
8	-0.006895793	-0.011758776
9	-0.009457720	0.016253318

Table 3: Random Effect: Item Level Variation

Experiment	Intercept (OBJ)	Extraction[T.SUBJ]
Experiment 3	0.09950918	-0.21479633
Experiment 5	-0.02671837	0.05767315
Experiment 4	-0.02396478	0.05172936
Experiment 6	-0.06004369	0.12960779
Experiment 2	0.01121766	-0.02421396

 Table 4: Random Effects: Experiment Variation