A slightly different version appeared in Yelena Fainleib, Nicholas LaCara, Yangsook Park (eds), *Proceedings of the 41th Annual Meeting of the North East Linguistic Society (NELS 41) – 2010*, pp. 233–246. Amherst, MA: Graduate Linguistic Student Association (GLSA). 2013.

Isolating Processing Factors in Negative Island Contexts[*]

Simone Gieselman, Robert Kluender & Ivano Caponigro

University of California, San Diego

1. Introduction

The phenomenon referred to as *negative islands* was originally observed by Ross (1984): while negation does not interfere with the extraction of arguments (1a), the extraction of adjuncts over negation renders sentences unacceptable (1b).

(1) a. Which project didn't the intern complete __ conscientiously?
b. *How didn't the intern complete the project __?

A number of proposals have been made in the theoretical linguistics literature that account for the difference between (1a) and (1b) in terms of global constraints operating within the grammar. Building on previous findings in the psycholinguistics literature, we used acceptability judgment measures to provide a new window into our understanding of negative islands. On the basis of results from two such studies, we argue that negative islands are not a unitary phenomenon due to a single global grammatical constraint, but rather the by-product of the simultaneous co-occurrence of different processing factors.

The paper is structured as follows. In section 2, we show that alongside the global constraints proposed in existing accounts of negative islands, there is abundant evidence in the psycholinguistics literature that each of the individual factors that figure into negative islands – namely negation, extraction, and referentiality – incurs its own processing cost. The results from the acceptability judgment studies reported in section 3 demonstrate the importance of taking these individual factors into consideration when analyzing negative islands. In Experiment 1, we investigate the effects of the factors negation and extraction, and in Experiment 2 we additionally manipulate the factor of referentiality. In section 4, we discuss the results of these experiments and suggest

[*] Thanks to Jonathan Cohen, Donka Farkas, Grant Goodall, Andrew Kehler, Daniel Lassiter, Florian Schwarz, Jon Sprouse and Satoshi Tomioka for helpful discussion, and to Arlene Aguilar, Masha Sokolov, and Tabitha Winders for help with stimulus preparation; this paper was supported in part by the National Institute on Deafness and Other Communication Disorders, award # T32 DC000041.

possible interpretations and consequences. Section 5 ends with our conclusions.

2. Grammatical Constraints and Processing Factors

We begin by discussing two prominent proposals that posit global constraints on negative islands within the grammar. We then briefly review evidence from the psycholinguistics literature showing that the factors known to play a role in negative islands (i.e., negation, extraction and referentiality) are all associated with independent processing costs. This suggests that the perception of negative islands may be due to the co-occurrence of these different processing factors rather than to a global grammatical constraint.

The best-known syntactic explanation for the contrast in (1a-b) is likely that of Rizzi (1990; 1992). Rizzi proposes that negation is a selective barrier to extraction and that only referential (but not non-referential) expressions can escape this barrier. Referential expressions (e.g. *which project*) differ from non-referential expressions (e.g. *how*) in bearing θ-roles and are therefore, according to Rizzi, assigned referential indices. Non-referential expressions leave traces without indices. Under the assumption that the properties of negation justify its syntactic positioning in the A'-specifier position of TP, negation can act as a potential antecedent governor for the trace of the extracted *wh*-phrase. Since the principle of *relativized minimality* rules out a syntactic configuration in which a potential antecedent governor (here: negation) interrupts the antecedent government relationship between an extracted element and a non-referential trace, sentences featuring non-referential argument extraction over negation are ruled out by the syntax. Negation does not interrupt the binding relationship between an extracted element and a referential trace, rendering sentences such as (1a) grammatical. Syntactic accounts of negative islands have however been challenged by the observation that semantic factors like discourse-linking (Pesetsky 1987), the iterability of the action denoted by the predicate (Szabolcsi & Zwarts 1993) and the presence of modal verbs (Kuno & Takami 1997; Fox & Hackl 2006; Abrusán 2008) affect the acceptability of sentences featuring extraction over negation. As a result, negative islands have been subjected to alternative semantic analyses, the most influential being that of Szabolcsi & Zwarts (1993).

According to Szabolcsi & Zwarts (1993), the acceptability of sentences featuring extraction over negation depends on the denotation domain of the extracted constituent. Extracted constituents denoting finite sets of individuals can take scope over negation because the Boolean operation required by negation is defined in this case (i.e. given a set of individuals, it is always possible to compute the complement set with respect to the domain of individuals D in the model). Thus referential arguments such as *which*+NP phrases can scope over negation (1a). However, the complement set cannot be computed in the case of expressions that denote a set whose elements are partially ordered, and therefore amount phrases like *how-much*+NP cannot scope over negation (1b).

In sum, what these theoretical linguistic accounts have in common is that they explain the perception of unacceptability in the case of negative island violations with global (syntactic or semantic) principles of grammar. Our approach to negative islands is based on a different idea, first proposed by J.D. Fodor (1983) and Ross (1987). These two

independent proposals share the intuition that the perception of unacceptability on a global level – both generally as well as more specifically within island contexts – is related to costs that accumulate across the course of a sentence. Here we pursue the hypothesis that the cumulative processing costs associated with negation, extraction and (non-)referentiality contribute to the perception of unacceptability in negative island contexts. This hypothesis further predicts that combinations of these factors should yield gradient effects of acceptability. Significantly, each of these three factors has a demonstrated processing cost associated with it in the psycholinguistics literature.

Over the past five decades, the processing difficulty associated with negation has been shown using a diverse range of experimental techniques other than acceptability judgments. The cost of negation is reflected in higher error rates (Wason 1961), longer response times (Slobin 1966), greater cortical activation (Carpenter et al. 1999) and larger brain responses (Staab 2007) for simple negative sentences compared to corresponding affirmative sentences. It has also been demonstrated that discourse context is important for the processing of negation (Wason 1965; Staab 2007; Nieuwland & Kuperberg 2008).

Evidence for the processing cost associated with extraction has nearly as long of a history as that for negation (Ford 1983; Frazier 1987; Kluender & Kutas 1993a; Fiebach et al. 2002). This cost is most commonly demonstrated in longer reading times for object vs. subject relative clauses (King & Just 1991). Importantly, none of these experimental manipulations featured extraction over an operator or barrier. The processing of object vs. subject relative clauses also elicits greater cortical activation (Just et al. 1996; Stromswold et al. 1996) and larger brain responses (King & Kutas 1995).

Turning to referentiality, both the theoretical linguistics (Kroch 1989; Rizzi 1990; 1992) and psycholinguistics literature have noted its effect on extraction, as reflected in acceptability and reading time data (Warren & Gibson 2002; Hofmeister 2007; Sag et al. 2007). However, the term 'referentiality' is not as clear-cut as the other two factors that figure in negative islands; there is some disagreement as to which types of constituents are referential or non-referential expressions. While some theoretical linguists for instance categorize *how-many*+NP as non-referential (Rizzi 1990; Chung 1994), others argue that *how-many*+NP is ambiguous between an amount (non-referential) reading and an individual (referential) reading (Cinque 1990; Szabolcsi & Zwarts 1993).

In sum, the psycholinguistics literature provides evidence that three factors that seem to play a role in negative island contexts – negation, extraction and referentiality – all have associated processing costs. Existing theoretical linguistic accounts primarily focus on explaining negative island effects in terms of global grammatical constraints. In two acceptability rating studies, our aim was to identify and isolate the various building blocks that create the perception of unacceptability in negative island contexts.

3. Acceptability Rating Studies

We experimentally tested our hypothesis that the perception of ungrammaticality in the case of negative island violations is due to the individual and cumulative processing costs

of the factors negation, extraction and referentiality. We examined the extent to which the costs of these three factors affected acceptability ratings individually and in combination.

3.1 Negation and Extraction

Like adjunct extraction (1b), argument extraction (1a) involves both negation and extraction. If the phenomenon of negative islands is at least partly attributable to such processing factors, an acceptability judgment study should reveal a drop in acceptability when an argument is extracted over negation (1a), even in cases in which the sentence is predicted not to violate any grammatical constraint and is pragmatically plausible.

3.1.1 Methods

Participants
28 undergraduate students from the University of California, San Diego who were native speakers of English received course credit for their participation.

Materials
24 sets of six parallel questions were constructed. Each set contained positive (2a) and negative (2b) yes/no-questions, positive (2c) and negative (2d) subject *wh*-questions and positive (2e) and negative (2f) object *wh*-questions. Yes/no-questions were included because they involve no extraction and therefore reveal pure effects of negation, subject *wh*-questions because they involve *wh*-quantification but not extraction over negation, and object *wh*-questions because they do require extraction over negation. All questions had the same basic structure: a subject consisting of a determiner and a noun, an object consisting of some type of determiner and a noun, a main verb and an adjunct consisting of a preposition, a determiner and a noun, as exemplified by the item set in (2). Item sets differed in their lexical realization such that every item set had a different verb, subject and object, but two of the 21 final prepositional phrases were repeated.

(2) a. Did the politician support the bill in the caucus?
b. Didn't the politician support the bill in the caucus?
c. Which politician supported the bill in the caucus?
d. Which politician didn't support the bill in the caucus?
e. Which bill did the politician support in the caucus?
f. Which bill didn't the politician support in the caucus?

All questions in all item sets and conditions were designed to be acceptable and felicitous and to avoid known grammatical violations. The felicity of condition (2f) with extraction over negation was particularly important because we were interested not in how felicity might affect acceptability ratings, but in how known processing costs of negation and extraction would.

With respect to the realization of negation, we decided to use the clitic *n't* attached to the auxiliary verb instead of the independent morpheme *not* directly preceding the main verb. There were two reasons for this choice. First, except for (2c-d), we were

thereby able to avoid a confound of sentence length due to the addition of an extra word in negative vs. positive sentences. Second, we wanted our experimental stimuli to be as acceptable as possible, and a pilot study had shown that undergraduate UCSD students preferred yes/no-questions with negation attached to the auxiliary (*Didn't the intern complete her project?*) over yes/no-questions with negation realized as an independent morpheme preceding the main verb (*Did the intern not complete her project?*).

Experimental sentences were supplemented by 18 positive and 18 negative filler sentences; the ratio of positive to negative sentences across the entire set of stimuli was thus 1:1. Filler sentences consisted of both yes/no- and *wh*-questions in order to create equal numbers of each type across the entire stimulus set. Since the experimental sentences contained no known violations of grammaticality, 30 of the 36 fillers did, yielding equal numbers of acceptable and unacceptable sentences in the experiment. Fillers ranged over a wide variety of levels of unacceptability, including subject-verb agreement violations, tense violations and island violations.

Experimental stimuli were rotated through a Latin square design to form six lists, such that every participant saw only one item per set and four different lexicalizations of each condition. Each list contained 24 experimental items with 36 randomly interspersed filler sentences. As each list was also presented in reverse order, there was a total of 12 lists, and each participant was randomly assigned to one of these.

Procedure

Participants were run in 15-minute sessions on a computer in the Language Lab of the UCSD Linguistics Department. After giving informed consent, participants were instructed to use their intuitions as native speakers of English to judge the naturalness of sentences presented to them one at a time on the computer screen. Underneath each sentence on the screen was a series of buttons numbered one to seven. Participants were instructed to click on one of the lower numbers for 'bad' sentences and one of the higher numbers for 'good' sentences, according to their judgment. There was no time pressure.

Analysis

Acceptability ratings were subjected to a repeated measures analysis of variance (ANOVA) with two within-group factors, NEGATION (positive vs. negative) and QUESTION TYPE (yes/no-question vs. subject *wh*-question vs. object *wh*-question) The Tukey HSD method was used to compute subsequent multiple pair-wise comparisons.

Predictions

The goal of this first experiment was to investigate how negation and extraction affect acceptability by themselves, in the absence of the factor (non-)referentiality. We therefore compared *wh*-questions in which the *wh*-constituent was always referential; when the *wh*-phrase was the object of the verb, it was extracted over negation (2f). While the global grammatical constraints discussed above do not predict any type of violation in such configurations, a processing account predicts that the individual processing costs incurred by negation and extraction should be reflected in the acceptability ratings.

3.1.2 Results

Results are presented in Figure 1 based on the mean acceptability ratings and standard deviations shown in Table 1. Note that all negative questions received numerically lower ratings than corresponding positive questions. This contrast was clearly largest for the object *wh*-questions, where it was also significant.

	Yes/No-Questions	Subject *Wh*-Questions	Object *Wh*-Questions
positive	6.7 (0.8)	6.3 (1.1)	6.5 (1.0)
negative	**6.2** (1.2)	**5.8** (1.4)	**4.8** (1.6)

Table 1: *Mean acceptability ratings (and standard deviations) from Experiment 1*

A repeated measures ANOVA revealed a main effect of NEGATION [$F1(1,27) = 67.116$, $p < 0.001$, $\eta^2_p = 0.71$; $F2(1,23) = 71.238$, $p < 0.001$, $\eta^2_p = 0.75$] driven not only by robust pairwise differences between positive (2e) and negative (2f) object *wh*-questions ($p < 0.001$), but also by robust differences of negative subject *wh*-questions (2d) from positive yes/no-questions (2a; $p = 0.001$) on the one hand and from positive object *wh*-questions (2e; $p < 0.05$) on the other. A main effect of QUESTION TYPE [$F1(2,54) = 24.744$, $p < 0.001$, $\eta^2_p = 0.47$; $F2(2,46) = 16.687$, $p < 0.001$, $\eta^2_p = 0.42$] was due to significantly lower ratings for negative object *wh*-questions (2f) in comparison to both negative yes/no- (2b; $p < 0.001$) and negative subject *wh*-questions (2d; $p < 0.001$). There was also an interaction of NEGATION x QUESTION TYPE [$F1(2,54) = 17.015$, $p < 0.001$, $\eta^2_p = 0.38$; $F2(2,46) = 9.6163$, $p < 0.001$, $\eta^2_p = 0.29$] reflecting the fact that negative object *wh*-questions (2f) were rated lower than all other positive and negative question types (all $p \leq 0.001$). Additional analysis and a follow-up study confirmed that reported effects were not due to the animacy or the singular/plural nature of the extracted object.

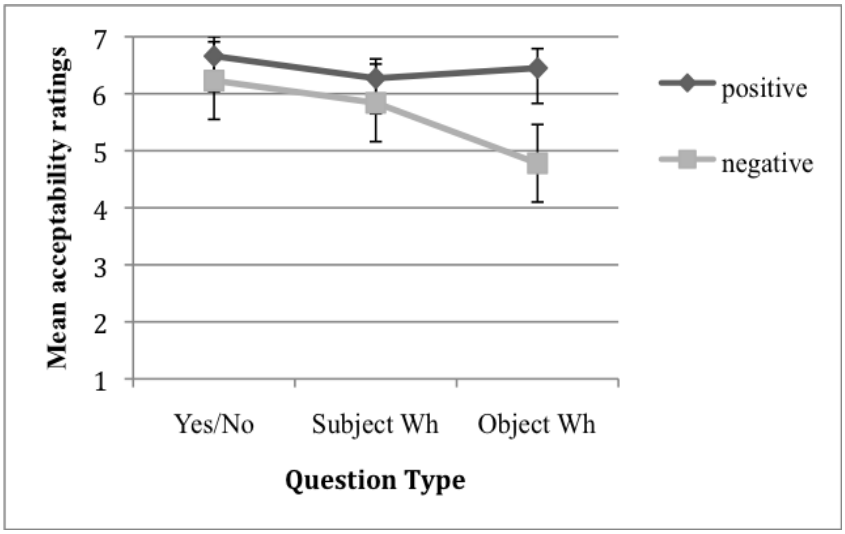

Figure 1: *Mean acceptability ratings from Experiment 1*

3.1.3 Summary

Negative object *wh*-questions behaved as predicted – based on independently established processing costs for negation and extraction – in precipitating a substantial drop in acceptability relative to other question types. In pairwise comparisons, the presence of negation caused smaller but nonetheless significant drops in acceptability independent of extraction. These results point to the fact that (i) the presence of negation impacts sentence acceptability and (ii) extraction of referential object arguments over negation significantly lowers acceptability ratings relative to other question types, even when no grammatical principles or constraints are violated. Note in particular that while the drop in acceptability for the extraction of referential objects over negation was highly significant, its acceptability rating of 4.8 still fell in the upper range of the 7-point scale.

3.2 Negation, Extraction and (Non-)Referentiality

Experiment 1 showed that the costs of negation and extraction lower acceptability ratings even when a referential argument is extracted. By hypothesis, the extraction of a non-referential argument should lead to a further decrement in acceptability over and above the decrements caused by negation and extraction observed in Experiment 1.

3.2.1 Method

Participants
28 different undergraduate students from the University of California, San Diego who were native speakers of English received course credit for their participation.

Materials
30 sets of 10 parallel questions were constructed, as shown in (3). Questions were presented in both positive and negative variants to demonstrate the effects of negation on acceptability ratings for all question types. As in Experiment 1, both subject (3c-f) and object (3g-j) *wh*-questions were included to show the effects of extraction. However, in this experiment, subject and object *wh*-questions with both *which*+NP (3c-d, 3g-h) and *how-many*+NP (3e-f, 3i-j) phrases were used in order to highlight the effects of referentiality on acceptability ratings. *How-many*+NP phrases lie somewhere on the continuum between fully referential *wh*-phrases like *which*+NP and non-referential *wh*-phrases like *how*, *how-much*, *how*+AdjP or *how*+AdvP. According to Szabolsci & Zwarts (1993), *how-many*+NP *wh*-phrases are potentially ambiguous between an amount (non-referential) and an individual (referential) reading (cf. also Cinque 1990). We chose *how-many*+NP for our non-referential conditions because it is syntactically similar to *which*+NP and can be constructed from the same set of head nouns. Moreover, the use of *how-many*+NP phrases allowed us to create experimental questions for which it was relatively easy to imagine a felicitous supporting context, thereby limiting the role of plausibility as a factor in any ensuing drops in acceptability.

(3) a. Did the interns complete their projects during the internship?
b. Didn't the interns complete their projects during the internship?
c. Which interns completed their projects during the internship?
d. Which interns didn't complete their projects during the internship?
e. How many interns completed their projects during the internship?
f. How many interns didn't complete their projects during the internship?
g. Which projects did the interns complete during the internship?
h. Which projects didn't the interns complete during the internship?
i. How many projects did the interns complete during the internship?
j. How many projects didn't the interns complete during the internship?

Experimental sentences were supplemented by 60 filler sentences according to the same criteria used in Experiment 1. Stimuli were then rotated through a Latin-square design to form ten lists, such that every participant saw only one item per set and three different lexicalizations of each condition. Each list contained 30 experimental items, with 60 filler sentences randomly interspersed among the experimental stimuli. Each of the ten lists of 90 sentences was also presented in reverse order such that the design included 20 lists total. Each participant was randomly assigned to one of the 20 lists.

Procedure
The same procedure as in Experiment 1 was used for Experiment 2.

Analysis
Acceptability ratings were subjected to a repeated measures ANOVA with three within-group factors: NEGATION (positive vs. negative), EXTRACTION (subject vs. object *wh*-question), and REFERENTIALITY (*which*+NP vs. *how-many*+NP). Yes/no-questions were included in the design to provide a baseline for purposes of comparison but not in the ANOVA, since by definition yes/no-questions are neither referential nor non-referential.

Predictions
The aim of this experiment was to see whether adding the factor (non-)referentiality to a design that already included negation and extraction would lead to a further drop in acceptability of otherwise plausible sentences. Neither semantic nor syntactic global constraints predict grammatical violations in the absence of extraction (3d&f). Accounts that assume that *how-many*+NP is non-referential (e.g. Rizzi 1990) predict a grammatical violation in the case of (3j), in which an argument of this type is extracted over negation. Processing considerations predict drops in acceptability both with and without extraction, with the largest drop occurring when all three factors are combined.

3.2.2 Results

Results are presented in Figure 2 based on the mean acceptability ratings and standard deviations shown in Table 2. As in Experiment 1, negative questions (3d, 3f, 3h, 3j) were rated less acceptable overall than positive questions (3c, 3e, 3g, 3i) [main effect of NEGATION: $F1(1,27) = 124.23$, $p < 0.001$, $\eta^2_p = 0.82$; $F2(1,29) = 186.62$, $p < 0.001$, $\eta^2_p = 0.86$]. Pairwise comparisons confirmed that this was due to significant differences

between all positive and negative *wh*-question pairs except subject *which*+NP (3c&d) questions [positive (3e) vs. negative (3f) subject *how-many*+NP questions: $p = 0.006$; positive (3g) vs. negative (3h) object *which*+NP questions: $p < 0.001$); positive (3i) vs. negative (3j) object *how-many*+NP questions: $p < 0.001$].

	Yes/No Baseline	*Which*+NP Subject *Wh*	*How many*+NP Subject *Wh*	*Which*+NP Object *Wh*	*How many*+NP Object *Wh*
positive	6.5 (1.0)	6.5 (0.9)	6.6 (1.0)	6.4 (0.9)	6.5 (1.1)
negative	**6** (1.6)	**5.8** (1.4)	**5.4** (1.6)	**4.7** (1.7)	**3.5** (1.7)

Table 2: *Mean acceptability ratings (and standard deviations) from Experiment 2*

Object *wh*-questions (3g-j) were likewise rated less acceptable overall than corresponding subject *wh*-questions (3c-f) [main effect of EXTRACTION: $F1(1,27) = 30.5$, $p < 0.001$, $\eta^2_p = 0.53$; $F2(1,29) = 34$, $p < 0.001$, $\eta^2_p = 0.54$]. An interaction of NEGATION x EXTRACTION [$F_1(1,27) = 41.285$, $p < 0.001$, $\eta^2_p = 0.6$; $F_2(1,29) = 43.996$, $p < 0.001$, $\eta^2_p = 0.63$] reflected the fact that differences between corresponding subject and object *wh*-questions were significant only in the presence of negation [negative subject (3d) vs. object (3h) *which*+NP questions: $p = 0.001$; negative subject (3f) vs. object (3j) *how-many*+NP questions: $p < 0.001$]; corresponding positive *wh*-subject and *wh*-object question pairs (3c&3g, 3e&3i) did not differ significantly from each other. As in Experiment 1, negative object *which*+NP questions (3h) were rated significantly lower in pairwise comparisons including yes/no-questions (3a&b) than all other question types (all $p \leq 0.001$), except for negative subject (3f) and object (3j) *how-many*+NP questions.

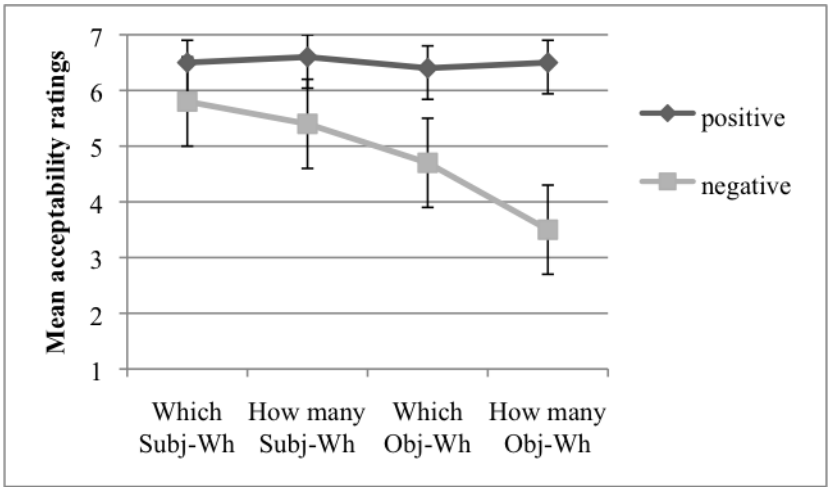

Figure 2: *Mean acceptability ratings from Experiment 2*

Crucially for our hypothesis, *how-many*+NP questions (3e-f, 3i-j) were rated less acceptable overall than *which*+NP questions (3c-d, 3g-h) [main effect of REFERENTIALITY: $F1(1,27) = 18.1$, $p < 0.001$, $\eta^2_p = 0.39$; $F2(1,29) = 8.01$, $p = 0.0084$, $\eta^2_p = 0.22$]. There was an interaction of NEGATION x REFERENTIALITY [$F_1(1,27) = 19.365$, $p < 0.001$, $\eta^2_p = 0.42$; $F_2(1,29) = 22.87$, $p < 0.001$, $\eta^2_p = 0.45$] attributable to a sizeable difference between negative object *which*+NP questions (3h) and negative object

how-many+NP questions (3j; $p = 0.001$) as well as to lower ratings for negative subject *how- many*+NP questions (3f) than for all positive question types (all $p \leq 0.02$). A three-way interaction of NEGATION x EXTRACTION x REFERENTIALITY [$F_1(1,27) = 14.75$, $p < 0.001$, $\eta^2_p = 0.38$; $F_2(1,29) = 4.3$, $p = 0.04$, $\eta^2_p = 0.12$] was due to significantly lower ratings of all negative *how-many*+NP questions (3f&j) in comparison to all other negative and positive question types ($p \leq 0.001$).

3.2.3 Discussion

Negative object *how-many*+NP (i.e. non-referential) questions led to a further drop in acceptability compared to negative object *which*+NP (i.e. referential) questions. Additionally, we found that various combinations of the three factors we manipulated – negation, extraction and referentiality – led to significant drops in acceptability, even in cases where only two out of three factors were simultaneously present. If we take as a baseline value the mean acceptability rating (6.5) for positive yes/no-questions (3a) and then step by step add in different factors, it is not the case that we see an effect on acceptability only when all factors co-occur. Adding negation to a subject *wh*-question with a non-extracted referential *wh*-phrase (3c vs. 3d) numerically decreases acceptability by more than half a point; if this referential *wh*-subject is made non-referential (3f) the acceptability drops further to 5.4; if a referential object *wh*-phrase is extracted over negation (3h: extraction and negation are manipulated) the acceptability drops to 4.7; and manipulating all three factors simultaneously (3j) further lowers acceptability to 3.5. This stepwise decrease in acceptability is consistent with Fodor's (1983) and Ross's (1987) suggestions that the costs of such factors are subtle but cumulative.

Note however that cumulative costs of extraction and referentiality show up only in the presence of negation. Negation causes a numerical drop in acceptability for all question types, while extraction and (lack of) referentiality lower acceptability only in combination with negation. We believe this reflects the relative strength of the negation effect, which is also visible in the partial eta squared value for the factor NEGATION. The partial eta squared test can roughly be described as the ratio of variance accounted for by an effect (e.g. Brown 2008), and NEGATION alone accounts for the greatest amount of variance in our data: 82% compared to somewhere between 12% and 60% for all other effects. In other work (Gieselman et al. *in press*) we have provided evidence that the heavy processing cost associated with negation may be related to the discourse conditions it imposes. This is consistent with other studies that point to the importance of discourse context in understanding the processing of negation (Staab 2007; Nieuwland & Kuperberg 2008). Since independent effects of extraction (Kluender & Kutas 1993a) and of referentiality in the context of extraction (Warren & Gibson 2002; Hofmeister 2007; Sag et al. 2007) have been demonstrated using on-line measures, the lack of isolated effects of extraction and referentiality on acceptability ratings in our study may be an artifact of the use of an off-line measure. Another reason may be that processing costs of extraction are known to be sensitive to embedding, and at times to sheer length, and all of our stimuli were short, simple, monoclausal structures.

4. General Discussion

In these two acceptability judgment studies, we investigated the effects of three factors that play a prominent role in negative island configurations: negation, extraction and referentiality. In Experiment 1, even without manipulating referentiality, we found that extraction over negation lowers acceptability ratings. Experiment 2 showed that, even in the absence of extraction, (non-)referentiality causes a drop in acceptability. Across both experiments, negation proved to be the most important factor in negative islands, as costs of extraction and (non-)referentiality surfaced only in the presence of negation.

One specific finding requires special attention: acceptability drops in the case of negative object *how-many*+NP questions led to a super-additive interaction. In other words, the drop in acceptability for negative object *how-many*+NP questions was greater than the cumulative effect of each individual factor (negation, extraction and referentiality) added together. On the assumption that sentences with acceptability ratings falling below four on a scale of one to seven are categorically ungrammatical, one could simply argue that negative object *how-many*+NP questions with a mean acceptability of 3.5 fall outside the range of the competence grammar. It is then tempting to attribute the three-way super-additive interaction to the triggering of a global grammatical constraint. We think that this would be the wrong interpretation, however, for the following reasons.

The three-way interaction was not the only super-additive interaction in our results. There was a two-way super-additive interaction between negation and extraction in Experiment 1 that was replicated in Experiment 2. There are no obvious semantic or syntactic reasons in existing accounts of negative islands for negation to interfere in a super-additive fashion with the extraction of individuated/referential *which*+NP arguments in this way. Even if such reasons could be constructed, existing accounts cannot accommodate the fact that the presupposition trigger *also* interferes with the extraction of *which*+NP arguments in the same super-additive way as negation (and to the same degree, with a mean acceptability rating of 4.8), while the adverb *just* – which unlike negation imposes no conditions on the discourse context – exhibits no such effects (Gieselman et al. *in press*). Additionally, Experiment 2 revealed a super-additive interaction between negation and referentiality in the absence of extraction. While it may be possible to derive an interaction of this sort from existing semantic accounts of negative islands (Szabolcsi and Zwarts 1993), it is generally not predicted under traditional syntactic accounts of island phenomena in the complete absence of any type of movement over negation. In view of these unexpected super-additive two-way interactions, the fact that there is a super-additive three-way interaction when all three factors are combined seems less surprising. In other words, while we cannot definitively rule out the logical possibility of a grammatical constraint applying when all three factors are combined, it is certainly not the only or even the most plausible account of the data.

With regard to referentiality, Kluender (1998) predicted that the more referential expressions are, the better they are anchored in working memory and the easier the retrieval. Experimental results from Hofmeister (2007) among others support this view. One might translate this into Lewis & Vasishth's (2006) model of working memory in

terms of activation levels: the higher the initial activation level of an extracted constituent in working memory, the easier the retrieval process for purposes of integration later on.

In what follows, we outline potential future directions for a processing account based on these findings. By hypothesis, the interaction between negation and extraction might be due to the fact that negation and extraction draw on related cognitive resources.[1] This idea is consistent with the event-related brain potential (ERP) literature on negation and extraction, in which negation (Staab 2007; Luedtke et al. 2008) and extraction (King & Kutas 1995; Fiebach et al. 2002; Felser et al. 2003; Phillips et al. 2005) elicit similar brain responses. Effects of extraction have previously been related to holding an extracted constituent in syntactic working memory[2] until it can be integrated into the thematic structure of the sentence (Gibson 1991; Kluender & Kutas 1993a; King & Kutas 1995). In view of ERP evidence that context plays a crucial role in on-line integration of negation into the truth conditions of a sentence (Staab 2007; Nieuwland & Kuperberg 2008), we hypothesize that the processing cost of negation is due to the working memory costs of context retrieval (Gieselman et al. *in press*). This predicts that other elements that impose discourse conditions should elicit effects similar to those elicited by negation, as in fact turned out to be the case with the presupposition trigger *also* (Gieselman et al. *in press*).

5. Conclusion

In this study we investigated fine-grained details of negative islands. Specifically, we identified the individual and cumulative effects of various factors (negation, extraction and referentiality) that contribute to the perception of what is known as a negative island. Taken together, these factors predicted the gradience in acceptability found in our data. Interestingly, previous studies of other types of weak islands (e.g. wh-islands, Kluender & Kutas 1993b) showed similar gradient effects. This suggests that there may be commonalities among the factors that are at play in various weak islands.

References

Abrusán, Márta. 2008. A Semantic analysis of negative islands with manner questions. In Proceedings of Sinn und Bedeutung, 12: 1 – 16. Grønn, Atle (ed.).

Brown, James Dean. 2008. Effect size and eta squared. JALT Testing & Evaluation Newsletter, 12: 38-43.

[1] We say "related" and not "equivalent" because the data from the current acceptability judgment experiments do not support the idea that negation and extraction produce the same effects: negation is reflected in acceptability ratings in the absence of extraction but not vice versa.

[2] Sprouse et al. (2009) showed that there are no correlations between working memory capacity as measured with a serial recall task or *n*-back test and the perception of island violations, and used this finding to argue against processing accounts of island phenomena. Note however that for both measures of working memory, the use of serial order information is required. According to the current working memory model of Lewis & Vasishth (2006), access to serial order information is too slow to support sentence comprehension. Therefore, it is expected that working memory as measured with tasks requiring serial order information would not necessarily correlate with on-line sentence comprehension.

Carpenter, P.A., Just, M.A., Keller, A.K., Eddy, W.F., and K.R. Thulborn. 1999. Time course of fMRI-activation in language and spatial networks during sentence comprehension. NeuroImage, 10: 216 – 224.

Chung, Sandra. 1994. Wh-agreement and "referentiality" in Chamorro. Linguistic Inquiry, 25: 1 – 44.

Cinque, Guglielmo. 1990. Types of A'-dependencies. MIT Press, Cambridge, MA.

Felser, C. Clahsen H and TF Münte. 2003. Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. Brain and Language, 87: 345 – 354.

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

Fodor, Janet Dean. 1983. Phrase structure parsing and the island constraints. Linguistics and Philosophy, 6: 163 – 223.

Ford, Marily. 1983. A method of obtaining measures of local parsing complexity throughout sentences. Journal of Verbal Learning and Verbal Behavior, 22.

Fox, Danny, and Martin Hackl. 2006. The universal density of measurement. Linguistics and Philosophy, 29: 537 – 586. Springer.

Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. Natural Language and Linguistic Theory, 5: 519 – 559.

Gibson, Edward. 1991. A computational theory of human linguistic processing: Memory limitations and processing breakdown. Doctoral Dissertation, Carnegie Mellon University, Pittsburgh, PA.

Gieselman, S., Kluender, R., and I. Caponigro. *In press*. Pragmatic processing factors in negative island contexts. In Proceedings of the Western Conference on Linguistics (WECOL), 22.

Hofmeister, Philip. 2007. Retrievability and gradience in filler-gap dependencies. Proceedings from the Annual Meeting of the Chicago Linguistic Society.

Just, Marcel, and Patricia Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99: 22 – 149.

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., and K.R. Thulborn. 1996. Brain activation modulated by sentence comprehension. Science, 274.

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

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

Kluender, Robert and Marta Kutas. 1993a. Bridging the gap: Evidence from ERPs on the processing of unbound dependencies. Journal of Cognitive Neuroscience, 5.

Kluender, Robert and Marta Kutas. 1993b. Subjacency as a processing phenomenon. Language and Cognitive Processes, 8: 573 – 633.

Kluender, Robert. 1998. On the distinction between strong and weak islands: a processing perspective, Syntax and Semantics, 29: 241 – 279.

Kroch, Anthony. 1989. Amount quantification, referentiality, and long wh-movement. Unpublished manuscript, University of Pennsylvania, Philadelphia.

Kuno, Susumu, and Ken-ichi Takami. 1997. Remarks on negative islands. Linguistic Inquiry, 28: 553 – 576.

Lewis, R., Vasishth, S., and J. Van Dyke. 2006. Computational principles of working memory in sentence comprehension, TRENDS in Cognitive Sciences, 10.

Lüdtke, J., Friedrich, C.K., De Filippis, M., and B. Kaup. 2008. Event-related potential correlates of negation in a sentence-picture verification paradigm. Journal of Cognitive Neuroscience, 20: 1355 – 1370.

Nieuwland, Malte and Gina Kuperberg. 2008. When the truth is not too hard to handle. Psychological Science, 19: 1213 – 1218.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In The representation of (in)definiteness, ed. Eric Reuland and Alice Meulen. MIT Press.

Phillips C., Kazanina N., and SH Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Research, 22: 407 – 428.

Rizzi, Luigi. 1990. Relativized Minimality. Cambridge, Mass.: MIT Press.

Rizzi, Luigi. 1992. Argument/adjunc(ta)symmetries, In Proceedings of NELS, 22.

Ross, John. 1984. Inner islands. In Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society. Berkeley Linguistics Society, UC Berkeley.

Ross, John. 1987. Islands and syntactic prototypes. Papers from the general session at the Twenty-Third Annual Reginal Meeting of the Chicago Linguistic Society, 1.

Slobin, D.I. 1966. Grammatical transformations and sentence comprehension in childhood and adulthood. Journal of Verbal Learning and Verbal Behavior, 5.

Staab, Jenny. 2007. Negation in context: electrophysiological and behavioral investigations of negation effects in discourse processing. Doctoral Dissertation, SDSU/UCSD.

Stromswold, K., Caplan, D., Alpert, N. and S. Rauch. 1996. Localization of syntactic comprehension by positron emission tomography. Brain and Language, 52.

Szabolcsi, Anna, and Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. Natural Language Semantics, 1: 235 – 284.

Sag, I., Hofmeister, P., and Neal Snider. 2007. Processing complexity in subjacency violations: The complex noun phrase constraint. In Proceedings from the Annual Meeting of the Chicago Linguistic Society, 43: 215 – 229.

Warren, Tessa, and Edward Gibson. 2002. The influence of referential processing on sentence complexity. Cognition 85: 79 – 112.

Wason, P.C. 1961. Response to affirmative and negative binary statements. British Journal of Psychology, 52: 133 – 142.

Wason, P.C. 1965. The contexts of plausible denial. In Journal of Verbal Learning and Verbal Behavior, 4: 7 – 11.

Simone Gieselman
Department of Linguistics, UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093
simone.gieselman@gmail.com

Robert Kluender
robert.kluender@gmail.com

Ivano Caponigro
ivano.caponigro@gmail.com