

- Rooij, Robert van. 2003. Questioning to resolve decision problems. *Linguistics and Philosophy* 26:727–763.
- Sharvit, Yael. 2002. Embedded questions and ‘de dicto’ readings. *Natural Language Semantics* 10:97–123.
- Szabolcsi, Anna. 1997. Background notions in lattice theory and generalized quantifiers. In *Ways of scope taking*, ed. by Anna Szabolcsi, 1–29. Dordrecht: Kluwer.

THE DIFFERENTIAL SENSITIVITY OF
ACCEPTABILITY JUDGMENTS TO
PROCESSING EFFECTS

Jon Sprouse

University of California, Irvine

1 Introduction

Linguists have agreed since at least Chomsky 1965 that acceptability judgments are too coarse-grained to distinguish between the effects of grammatical knowledge (what in Chomsky 1965 would be called competence effects) and the effects of implementing that knowledge (or performance effects). This granularity problem means that for any given putative grammatical phenomenon whose existence is demonstrated by acceptability judgments, it is logically possible that the unacceptability is an epiphenomenon of human language processing. To take a famous example, many works have argued that the island effects delineated in Ross 1967 are due to the processing burdens encountered at island boundaries, and not due to grammatical constraints (e.g., Kluender and Kutas 1993, Kluender 1998, 2004).

With the rise of refined experimental methodologies for collecting acceptability judgments, there has been a renewed interest in identifying the contribution of performance factors—in particular, processing factors—to acceptability judgments. For instance, Fanselow and Frisch (2006) report that local ambiguity in German can lead to increases in the acceptability of ultimately ungrammatical representations if the second possible representation is grammatical. Hofmeister et al. (2007) report that factors affecting the acceptability of Superiority violations also affect the processing of *wh*-questions as measured in reading times, suggesting that there might be a correlation between processing factors and the acceptability of Superiority violations.

While the picture that emerges from these studies is that acceptability judgments are affected by a wide range of processing effects, this squib presents two experiments that suggest that acceptability judgments are not affected by every processing effect. This differential sensitivity to processing effects suggests a potential evaluation metric for the plausibility of processing explanations: if the proposed processing effect exists independently of the structures under consideration,

I am grateful to Howard Lasnik, Norbert Hornstein, and Colin Phillips for many helpful conversations and for comments on an earlier draft. All remaining mistakes are my own.

it should be possible to show that the acceptability effect exists independently of the structure as well.

The experiments reported in this squib build upon one of the major findings of sentence-processing research: the active filling strategy. The active filling strategy is defined by Frazier and Flores d'Arcais (1989:332) as follows: "when a filler has been identified, rank the possibility of assigning it to a gap above all other options." In other words, the human parser prefers to complete long-distance dependencies as quickly as possible. Because the earliest completion site is not always the correct one, the active filling strategy entails the construction of many incorrect temporary representations. Just like their nontemporary counterparts, these temporary representations can be manipulated such that they are either completely grammatical, syntactically ungrammatical, or semantically implausible.

The experiments reported in this squib investigate whether these temporary representations affect the acceptability of the final representation, and if so, which type(s). The results suggest that syntactically ungrammatical temporary representations do lower the acceptability of the final representation, while semantically implausible temporary representations and completely grammatical temporary representations have no effect. This pattern of results suggests (a) that acceptability judgments interact with syntactic violations in a qualitatively different way than semantic violations and pure processing mistakes, and (b) that acceptability judgments are differentially sensitive to effects of sentence processing.

2 Syntactically and Semantically Ungrammatical Representations

In experiment 1, two paradigms that take advantage of the active filling strategy were taken directly from the sentence-processing literature to test the effects of syntactically ungrammatical and semantically ungrammatical temporary representations: the filled-gap paradigm (Crain and Fodor 1985, Stowe 1986) and the plausibility paradigm (Garnsey, Tanenhaus, and Chapman 1989, Tanenhaus, Carlson, and Trueswell 1989). The effects of both paradigms on reading times are so well established in the sentence-processing literature as to serve as standard tools for investigating online construction of filler-gap dependencies.

In the filled-gap paradigm, incremental construction of a *wh*-dependency proceeds until the processor encounters the first verb, at which point the active filling strategy mandates that the *wh*-dependency be completed. If the verb has an empty θ -position for the *wh*-filler, construction of the rest of the representation proceeds as usual as in (1a). If the verb has no empty θ -position, the dependency is still completed, but the following NP receives no θ -role. Thus, a θ -Criterion-violating temporary representation is created until an empty θ -position can be found for the *wh*-filler (in this case, a prepositional phrase), and the structure is reanalyzed as in (1b). In the sentence-processing literature, the effect

of this θ -Criterion-violating representation is manifested as slower reading times at the NP object of the first verb.

- (1) *The filled-gap paradigm: Gap and filled-gap conditions*
- a. My brother wanted to know *who* Ruth will bring _____ home to Mom at Christmas.
 - b. My brother wanted to know *who* Ruth will bring *us* home to _____ at Christmas.

In the plausibility paradigm, incremental construction of the *wh*-dependency again proceeds until the processor encounters the first verb, at which point the active filling strategy again mandates that the *wh*-dependency be completed. In this paradigm, the argument structure of the verb is not manipulated; instead, the plausibility of the *wh*-filler serving as an argument of the verb is manipulated. If the *wh*-filler is a plausible argument of the verb, construction of the remaining representation proceeds as usual as in (2a). However, if the *wh*-filler is an implausible argument of the verb, the completed dependency results in a semantically implausible temporary representation that persists until a plausible empty θ -position (in this case, a prepositional phrase) is encountered, and the structure is reanalyzed as in (2b). In the sentence-processing literature, the effect of this implausible representation is manifested as slower reading times at the first verb.

- (2) *The plausibility paradigm: Plausible filler and implausible filler conditions*
- a. John wondered *which general* the soldier killed _____ effectively and enthusiastically for _____ during the war in Korea.
 - b. John wondered *which country* the soldier killed _____ effectively and enthusiastically for _____ during the war in Korea.

2.1 Participants

Eighty-six University of Maryland undergraduates participated in experiment 1 for extra credit. All of the participants were self-reported native speakers of English. All were enrolled in an introductory linguistics course.¹

2.2 Materials and Design

The survey was 34 items long including practice items, and it took about 15 minutes to complete. Items for the filled-gap paradigm were reconstructed from the examples in Stowe 1986. Items for the plausibility paradigm were taken from Pickering and Traxler 2003. Each survey consisted of two tokens each of the conditions from the two

¹ Topics relevant to this study either had not yet been introduced in the course (such as *wh*-constructions and acceptability judgments) or were never introduced in the course (such as magnitude estimation and the active filling strategy).

paradigms (8 items), 4 acceptable fillers, 14 unacceptable fillers, and 8 practice items, to make up the 34-item total. Items were distributed among 24 lists using a Latin square distribution, and pseudorandomized such that no two conditions from the same paradigm were consecutive. The instructions were a modified version of the instructions distributed with the WebExp software suite (Keller et al. 1998). The reference sentence for both the practice and the experimental items was a three-clause declarative sentence containing a *whether*-island violation: *Mary figured out what her mother wondered whether she was hiding.*

2.3 Results

Results were divided by the reference score and log-transformed prior to analysis. The mean of the two tokens from each condition was obtained for each participant, and then paired *t*-tests were performed on the pairs of conditions for each paradigm. As table 1 indicates, there was a large and highly significant decrease in acceptability for filled gaps as compared to unfilled gaps, mirroring the direction of the filled-gap effect in the sentence-processing literature. However, there was no effect of implausibility. Even though there are no direct statistical comparisons across the groups such that the family-wise error rate need be corrected, it is clear that both of the significant *p* values are well under the conservative Bonferroni correction level of .0167. Furthermore, the trend in the direction of an implausibility effect is an order of magnitude weaker than the significant effect of filled gaps: power analyses reveal that the filled-gap effect reaches significance at 20 subjects, whereas the trend in the plausibility paradigm would require 400 subjects to reach significance.

2.4 Discussion

At first glance, the asymmetrical pattern of results from experiment 1 seems to suggest that temporary syntactic ungrammaticality affects

Table 1
Results and paired *t*-tests for experiment 1

Condition	Mean	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>r</i>
Long-distance	.08	.19				
Short-distance	.20	.17	85	5.3	.001	.50
Filled-gap	.03	.24				
Unfilled-gap	.16	.24	85	5.6	.001	.52
Implausible filler	.09	.18				
Plausible filler	.10	.20	85	0.5	.608	

global judgments, whereas temporary semantic ungrammaticality does not. Unfortunately, there is a second possible explanation: reanalysis. By definition, the filled-gap condition of the filled-gap paradigm involves abandoning one structure and constructing a second one, a type of syntactic reanalysis: after the association between the *wh*-filler and the verb occurs, subsequent integration of the NP object fails. The parser must then reanalyze the structure such that the *wh*-filler is then associated with the preposition. In other words, the parser integrates the filler twice. However, the gap condition of the paradigm involves no such reanalysis because there is no extra NP object. It could be the case, then, that the difference in acceptability between the two conditions in the filled-gap paradigm is an effect of reanalysis on the judgment. This would also account for the lack of effect in the plausibility conditions: in both conditions, the *wh*-filler is initially associated with the verb and later reanalyzed as the object of the preposition. If reanalysis leads to a decrease in acceptability, both conditions in the plausibility paradigm should decrease equally, and one would expect a significant effect only in the filled-gap paradigm. Experiment 2 was designed to tease apart these two hypotheses (asymmetry due to temporary unacceptability versus asymmetry due to reanalysis).

3 The Reanalysis Confound

By definition, there is no way to eliminate reanalysis from the filled-gap and plausibility paradigms. However, it is possible to add reanalysis to the gap condition of the filled-gap paradigm, thus making it completely parallel to the plausibility paradigm in that both conditions will involve reanalysis. If the asymmetry in the presence of reanalysis across the two paradigms was the source of the asymmetry in the results for experiment 1, then eliminating the reanalysis asymmetry should eliminate the asymmetry in the results such that both paradigms return no effect. Furthermore, by comparing the new gap + reanalysis condition to the original gap condition, it is possible to isolate the effect of reanalysis alone, if it exists. This comparison investigates the effect of a temporary grammatical representation on the judgment of the final representation—in other words, the effect of processing difficulty without ungrammaticality—setting up the three-way comparison of temporary representations discussed in section 1.

3.1 Participants

Twenty-one University of Maryland undergraduates participated in experiment 2. All were self-reported native speakers of English without any formal training in linguistics. All were paid for their participation.

3.2 Materials and Design

The materials for experiment 2 were adapted from the materials for the plausibility paradigm in experiment 1, which were themselves

adapted from Pickering and Traxler 2003. As mentioned above, three conditions were used to test whether the source of the asymmetry from experiment 1 was the reanalysis asymmetry:

(3) *Filled gap + reanalysis (FG+R)*

John wondered *which general* the soldier killed *the enemy* effectively and enthusiastically for ____ during the war in Korea.

(4) *Gap + reanalysis (G+R)*

John wondered *which general* the soldier killed ____ effectively and enthusiastically for ____ during the war in Korea.

(5) *Gap (G)*

John wondered *which general* the soldier killed ____ effectively and enthusiastically for our side during the war in Korea.

The competing hypotheses make different predictions: if reanalysis is the source of the asymmetry, then experiment 2 should yield no effect between FG+R and G+R because both conditions involve reanalysis, and a significant effect between G+R and G because there is an asymmetry in reanalysis; if the asymmetry is due to the nature of the representation constructed, then there should again be an effect between FG+R and G and also an effect between FG+R and G+R. This hypothesis makes no prediction about G+R and G, but that comparison will indicate whether reanalysis has any effect at all.

Eight lexicalizations of each triplet were constructed and distributed using a Latin square design. Each list contained 1 token of each condition. Ten additional conditions from an unrelated study were included as fillers (4 acceptable, 6 unacceptable). Eight practice items were also included, for a total of 21 items. The task was magnitude estimation, and the instructions were identical to those of experiment 1. The reference sentence was also identical.

3.3 Results and Discussion

As before, results were divided by the reference score and log-transformed prior to analysis. The results and *t*-tests for experiment 2 are presented in tables 2 and 3, respectively. Corroborating the results from

Table 2
Results for experiment 2

Condition	Mean	SD
Filled-gap	-.02	.22
Gap + reanalysis	.09	.22
Gap only	.11	.20

Table 3
t-tests for experiment 2

Condition 1	Condition 2	<i>df</i>	<i>t</i>	<i>p</i>	<i>r</i>
Filled-gap	Gap	20	2.8	.005	.53
Filled-gap	Gap + reanalysis	20	2.8	.005	.53
Gap	Gap + reanalysis	20	0.3	.37	

experiment 1, there was a large significant decrease in acceptability of the filled-gap condition compared to the gap condition. There was also a large significant decrease in acceptability of the filled-gap condition compared to the gap condition with reanalysis, and no effect between the gap condition and the gap condition with reanalysis. This is the pattern of results that was predicted above if reanalysis has no effect on acceptability, such that the presence of a filled-gap effect is due to the syntactically ungrammatical temporary representation.² Furthermore, the lack of effect between the two gap conditions suggests that reanalysis has no persistent effect on the judgment of the final representation; in other words, there is no judgment cost associated with abandoning one well-formed representation for another.³

4 The Differential Sensitivity of Acceptability to Processing Effects

At an empirical level, the results from the experiments reported in this squib reveal an asymmetry in the effects of temporary representations on global acceptability, suggesting that the judgment process treats syntactic difficulties in a qualitatively different way from semantic or processing difficulties. This seems to indicate that judgment tasks are tapping directly into syntactic knowledge in a very real sense. At a methodological level, these results demonstrate the sensitivity of formal judgment experiments: the ability to detect significant differences between two acceptable sentences opens the possibility of using judgment experiments to explore phenomena that are typically the domain of sentence-processing studies. And at a theoretical level, these results

² All of the *p* values were one-tailed. Both of the significant values were well below the Bonferroni corrected level of .0167, even at their two-tailed value of .01.

³ An anonymous reviewer observes that the “filled-gap effect” from experiment 1 could be a distance effect: the distance between the *wh*-filler and the gap site is longer in the filled-gap condition than the gap condition. This is also true of the two gap conditions in experiment 2: the distance between the *wh*-filler and the gap site is longer in the reanalysis condition than the gap condition. The fact that there was no significant difference between these two conditions in experiment 2 suggests that the effect in experiment 1 was not a distance effect.

indicate that some, but not all, processing effects affect acceptability judgments. This differential sensitivity suggests that it is possible to use acceptability judgments to investigate the predictions of processing-based analyses of acceptability facts by first determining whether the processing effects in question affect acceptability at all, and then whether the acceptability of theoretically related phenomena are similarly affected (or unaffected).

References

- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Crain, Stephen, and Janet Fodor. 1985. How can grammars help parsers? In *Natural language parsing: Psycholinguistic, computational, and theoretical approaches*, ed. by David Dowty, Lauri Karttunen, and Arnold Zwicky, 94–128. Cambridge: Cambridge University Press.
- Fanselow, Gisbert, and Stefan Frisch. 2006. Effects of processing difficulty on judgments of acceptability. In *Gradience in grammar*, ed. by Gisbert Fanselow, Caroline Féry, Mattias Schlesewsky, and Ralf Vogel, 291–316. Oxford: Oxford University Press.
- Frazier, Lyn, and Giovanni Flores d'Arcais. 1989. Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28:331–344.
- Garnsey, Susan, Michael Tanenhaus, and Rachel Chapman. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18:51–60.
- Hofmeister, Philip, T. Florian Jaeger, Inbal Arnon, Ivan Sag, and Neal Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in search of its evidential base*, ed. by Sam Featherston and Wolfgang Sternefeld, 185–206. Berlin: Mouton de Gruyter.
- Keller, Frank, Martin Corley, Steffan Corley, Lars Konieczny, and Amalia Todirascu. 1998. WebExp. Technical report hrc/tr-99, Human Communication Research Centre, University of Edinburgh.
- Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In *The limits of syntax*, ed. by Peter Culicover and Louise McNally, 241–279. Syntax and Semantics 29. San Diego, CA: Academic Press.
- Kluender, Robert. 2004. Are subject islands subject to a processing account? In *WCCFL 23*, ed. by Angelo J. Rodriguez, Vineeta Chand, Ann Kelleher, and Benjamin Scheiser, 475–499. Somerville, MA: Cascadilla Press.
- Kluender, Robert, and Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8:573–633.
- Pickering, Martin, and Michael Traxler. 2003. Evidence against the use of subcategorization frequency in the processing of unbounded

- dependencies. *Language and Cognitive Processes* 18:469–503.
- Ross, John Robert. 1967. Constraints on variables in syntax. Doctoral dissertation, MIT, Cambridge, MA.
- Stowe, Laurie. 1986. Parsing *wh*-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1:227–245.
- Tanenhaus, Michael, Greg Carlson, and John Trueswell. 1989. The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes* 4:211–234.