

Revisiting Satiation: Evidence for an equalization response strategy
Jon Sprouse

Abstract

This reply revisits the topic of syntactic satiation as first discussed in Snyder 2000. I argue that the satiation effect reported in Snyder 2000 is the result of a response strategy in which participants attempt to equalize the number of yes and no responses, a strategy enabled by the design features of Snyder's original experiment. Four predictions differentiate the response strategy from a true satiation effect. A series of nine experiments are presented to test these predictions. The results are discussed with respect to the nature of satiation, the stability of acceptability judgments, and the consequences for linguistic methodology.

Keywords: satiation, magnitude estimation, experimental syntax, island constraints, syntactic priming

1. Introduction

Linguists have long reported that some unacceptable sentences will begin to sound more acceptable after days or weeks of repeatedly judging their acceptability. While this has often been dismissed by linguists as a minor occupational hazard, critics of the use of acceptability judgments have sometimes used this fact as evidence of the instability of the data underlying syntactic theories. Snyder 2000 replies directly to these types of criticisms with the following logic: If instability is an inherent property of judgments, and not just a symptom of 'doing linguistics', then we should see it in non-linguists. Furthermore, general instability in judgments should manifest in every type of unacceptable sentence such that if only a subset of unacceptable sentences show the instability, then instability could be a property of specific structures or constraints, and not the judgment process. Calling the increase in acceptability *satiation*, Snyder demonstrated that undergraduates with no linguistics training demonstrate satiation for 3 out of 7 violation types tested. This provided evidence for his hypothesis, and raised the possibility of studying satiation as part of the study of syntactic theory. Thanks to Snyder, satiation became a new form of data for syntacticians to use in differentiating constraints

(or classes of constraints) rather than a liability.

Snyder 2000 also suggests that it is possible that satiating and non-satiating violations have entirely different sources within the language faculty. In particular, he suggests that satiating violations may be due to extra-grammatical factors, such as some sort of processing constraint or resource limitation - a suggestion that receives some support in a study by Luka and Barsalou (2005) on syntactic priming. Syntactic priming is the facilitation of a given structure through previous exposure to that structure. Facilitation of a structure can manifest as a higher probability of using that structure in speech (e.g., Bock 1986, Pickering & Branigan 1998), or as faster reading times (e.g., Noppeney & Price 2004, Kaschak & Glenberg 2004). Luka and Barsalou 2005 found that syntactic priming can even manifest as higher acceptability: exposure to structures in a reading task leads to higher acceptability in a rating task. This suggests that satiation might just be a token of syntactic priming. Under the assumption that the structure crucial to a putative grammatical violation must be represented in order to prime or be primed, the fact that some violations can prime and others cannot can be interpreted as evidence for two types of structures even within the class of sentences that linguistic theory says are grammatical violations: those that can be represented and those that cannot. This looks remarkably similar to the grammatical/ungrammatical distinction, and thus could suggest that some of the putative grammatical violations are not grammatical violations at all, but rather violations of other constraints on acceptability, such as ease of processing (see Sprouse 2007b for a discussion).

As this brief introduction indicates, satiation is a topic that touches upon the work of both linguists and psycholinguists, as it has implications for linguistic methodology, linguistic theory, and the nature of linguistic representations. Unfortunately, in the years since Snyder's original study, the results of satiation studies on unacceptable sentence types have yielded mixed results – a situation I will call the replication problem in section 2. In this paper, I re-examine satiation in light of the replication problem, suggesting instead that the satiation effects in Snyder 2000 may derive from a response strategy in which participants attempt to equalize the number of each response. I argue that this strategy is enabled by two design features of Snyder's original experiment: (i) the task offered only two response choices (yes and no), and (ii) the design included significantly more unacceptable sentences than acceptable sentences. A series of 9 experiments are presented to test four predictions that differentiate between an analysis in which satiation is an effect of the equalization strategy and an analysis in which it is an inherent property of violations as suggested in Snyder 2000. The results of these experiments suggest that the satiation effect may be a result of the equalization strategy. While this conclusion has the unfortunate consequence of neutralizing Snyder's suggestion that satiation may be a relevant form of data for syntactic theories, it also suggests that acceptability judgments are more robust than some critics have suggested (e.g. Edelman and Christiansen 2003).

2. The replication problem and the equalization strategy

2.1 Defining the replication problem

Snyder (2000) tested the following 7 violation types for a satiation effect:

Table 1: Violations tested in Snyder 2000

Adjunct Island	Who did John talk to Mary after seeing?
CNPC Island	Who does Mary believe the claim that John likes?
LBC violation	How many did John buy books?
Subject Island	What does John know that a bottle of fell off the table?
That-trace	Who does Mary think that likes John?
Want-for	Who does John want for Mary to meet?
Whether Island	Who does John wonder whether Mary likes?

He found significant satiation for CNPC Islands and Whether Islands, and marginal satiation ($p < .07$) for Subject Islands. To the best of my knowledge, since Snyder's original study there have been five comparable satiation studies (and one eye-tracking study, Braze 2002). These studies have differed only slightly from Snyder's original study, but have yielded very different results.¹ Postponing for a moment a detailed discussion of the Snyder 2000 design and definition of satiation effect, the results of these studies can be summarized in the following table to visually demonstrate the replication problem:

Table 2: Summary of satiation results and the replication problem

	Snyder 2000	Hiramatsu 2000	Goodall 2005	Sprouse 1	Sprouse 2	Sprouse 3
Adjunct						
CNPC	√		√			
LBC						
Subject	(√)	√				
That-trace		√				--
Want-for		√	--			--
Whether	√	√	--			

√ = significant effect, (√) = marginal effect, -- = not tested

Replicability can be a problem for any experimental result. However, Snyder's insight was that satiation may be a property inherent to certain violations and not others, based on the source of the violation within the language faculty. As such, replicability is crucial to the argument: if some violations satiate in some experiments, and others in other experiments, the effect stops looking like a property of violations, and more like a general property of the task. In the following section that is exactly what I propose satiation is, a task effect. However, it is not a property of all judgment tasks as assumed by critics of acceptability judgments: it is a response strategy to very specific design features of these experiments.

2.2. Enabling the equalization strategy

The experiment in Snyder 2000 was designed as follows: Participants were asked to judge sentences as acceptable or unacceptable by circling either *yes* or *no*. They saw 50

sentences (not including practice items and a post-test). The 50 sentences were split into 5 blocks of 10 sentences, with each block containing one of each of the 7 violations in Table 1, and 3 completely grammatical fillers. In other words, participants saw 35 unacceptable sentences and 15 acceptable sentences. The operational definition of satiation adopted by Snyder was as follows:

- (1) Operational definition of satiation (performed separately for each participant)
 1. Count the number of yes responses for each violation in the first two blocks
 2. Count the number of yes responses for each violation in the last two blocks
 3. If the number of yes's increased, the participant satiated on that violation; If the number of yes's decreased, the subject not-satiated on that violation.

This is where the equalization strategy comes into play. Let's assume for the sake of argument that participants are disconcerted by using one response, in this case *no*, 70% of the time, and prefer to achieve a 50/50 balance between *yes* and *no* responses. At the end of the third block (i.e., immediately before the crucial blocks in the analysis), participants will have seen 21 sentences that prompt *no* responses, and only 9 sentences that prompt *yes* responses – a 2:1 ratio. If they adopt the equalization strategy, they will respond by providing more *yes* responses in the final two blocks to attempt to equalize the ratio. In other words, they will respond by satiating according to the operational definition in (1).

2.3 Predictions of the yes-no strategy

The equalization strategy requires three components to converge: an unbalanced design, responses that can be tracked by participants, and the decision of participants to alter their responses based on this tracking. The satiation explanation, being a direct manifestation of the architecture of the language faculty, contrasts with this in requiring nothing more than the language faculty itself. By manipulating each of the requirements of the equalization strategy, the two explanations provide contrasting predictions:

1. *The Replication Problem*

The equalization strategy is a task-related response, while satiation is (presumably) a reflex of the architecture of the language faculty. A response strategy would predict the replication problem, as replication is predicated upon a significant number of participants employing the same strategy. A direct reflex of the language faculty is put in doubt by failures to replicate. Therefore the replication problem is evidence in favor of the equalization strategy.

2. *Limited response tasks and unbalanced designs*

The equalization response requires both an unbalanced design and the ability of participants to track their responses. Tracking is more likely with limited response tasks, such as the binary choice yes-no task. We've already seen that in at least some experiments of this type, judgment instability occurs. However, satiation also predicts this.

3. Unlimited response tasks and balanced designs

The equalization strategy predicts that eliminating the imbalance and inhibiting the ability of participants to track their responses will neutralize the satiation effect. The satiation explanation predicts that the effect should still exist.

4. Limited responses and re-balanced designs

By re-balancing the design, I mean shifting the balance to slightly favor a strategy in which *no* responses should increase over time: more grammatical items than ungrammatical items. Given a definition of satiation based on the number of *yes* responses, the equalization strategy predicts no satiation effect, but the satiation explanation predicts that it should still occur.

Since predictions 1 and 2 have already been discussed, the next two sections are devoted to reporting experiments that test predictions 3 and 4 respectively.³ As we shall see, the results favor the equalization strategy.

3. Unlimited response tasks and balanced designs

Magnitude estimation provides an ideal task for providing participants with a theoretically unlimited number of potential responses (Stevens 1957, Bard et al 1996). In linguistic magnitude estimation, participants are asked to rate the acceptability of a target sentence by using the acceptability of a reference sentence as a unit of measure. In Figure 1, the acceptability of the reference sentence is assigned an integer rating, in this case 100. The participant is then asked to rate the acceptability of the target item using the reference as a unit of measure. If the participant feels the target item is only half as acceptable as the reference, they would assign it a rating of 50. If it is twice as acceptable, they would rate it 200:

Figure 1: An example of linguistic magnitude estimation

Reference: What do you wonder whether Mary bought?
Acceptability: 100

Target Item: What did Lisa meet the man that bought?
Acceptability: ____

Because the positive number line is theoretically infinite, participants have an unlimited number of potential responses.

Five magnitude estimation experiments were created to test the prediction that satiation should not arise in experiments with unlimited responses and balanced designs. The first four experiments each tested a different type of island violation: Subject Islands, Adjunct Islands, Whether Islands, and CNPC Islands. These violations were chosen because Subject, Whether, and CNPC Islands had shown satiation in Snyder 2000, and because Adjunct Islands have been analyzed as theoretically similar to Subject Islands (e.g., Huang 1082). The fifth experiment tested CNPC Islands again, with the addition of context sentences to ensure that lack of context in the other experiments was not

influencing the results.

3.1 Design

The experiments were balanced according to the best practices of psycholinguistics (e.g., Kaan and Stowe ms): the ratio of acceptable sentences to unacceptable sentences was 1:1, and the ratio of fillers to target items was 2:1. To achieve this, the experiments employed a block design with each block containing: i) two tokens of the island violation, one ungrammatical filler (agreement violations in the Subject and Adjunct Island experiments, a type of Sentential Subject Island violation in the other experiments), and 3 grammatical fillers. The Subject and Adjunct Island experiments presented participants with 7 blocks for a total of 14 repetitions of the island violations. The other experiments presented participants with 5 blocks for a total of 10 repetitions of the island violations. For comparison, Snyder 2000 and Goodall 2005 presented 5 repetitions of each violation, and Hiramatsu 2000 presented 7 repetitions. The participants were University of Maryland undergraduates with no formal training in linguistics. The sample sizes were 20, 24, 20, 17, and 20 respectively (compared to 22 in Snyder 2000).⁵

3.2 Results

The results of each experiment were normalized divided by the value of the reference sentence and then log transformed following the standard practices of linguistic magnitude estimation (Bard et al 1996, Keller 2000) prior to analysis.⁶ Table 3 presents the mean judgment and standard deviation for each block in each experiment:

Table 3: Means and Standard Deviations for each block

	Subject		Adjunct		Whether		CNPC		CNPC w/c	
	M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)
2 repetitions	-.14	(.24)	-.28	(.36)	.08	(.34)	-.01	(.28)	-.03	(.45)
4 repetitions	-.15	(.21)	-.32	(.37)	.12	(.35)	.01	(.28)	.08	(.44)
6 repetitions	-.07	(.30)	-.38	(.45)	.16	(.38)	-.04	(.27)	.01	(.41)
8 repetitions	-.11	(.21)	-.27	(.30)	.14	(.34)	-.01	(.36)	.05	(.40)
10 repetitions	-.09	(.22)	-.21	(.39)	.16	(.30)	-.07	(.36)	.08	(.38)
12 repetitions	-.09	(.22)	-.33	(.34)						
14 repetitions	-.10	(.27)	-.27	(.47)						

Unlike the categorical distinction inherent in yes-no tasks, magnitude estimation responses are continuous. Translating the original satiation definition (an increase in yes responses) into a continuous measure is not entirely straightforward. If satiation is the crossing of the categorical boundary between yes and no, then there are two possible changes in the continuous judgment that could lead to crossing of the boundary. The first is obvious: the mean judgment could simply increase over time, and eventually cross the category boundary. The second is less obvious: the spread (variance) of the responses could increase such that some, but not all, of the responses cross the category boundary. Because the starting judgment of the violations is *no*, and because satiation is defined as

an increase in the number of yes responses, this could also yield a satiation effect in a yes-no judgment.

Since satiation in a magnitude estimation experiment can be defined as either a significant increase in mean acceptability or a significant increase in variance over the course of the experiment, repeated measures linear regressions were run on both the mean judgments and on the variances.⁷ And since the increases could manifest in fewer repetitions than the full number presented in these experiments, and thus could be neutralized by a single regression over the entire experiment, Table 4 reports the *p*-values for repeated measures linear regressions (following the procedure developed in Lorch and Myers 1990) after each block beginning with the second (effects are considered significant when the *p*-value is less than .05):

Table 4: Repeated measures linear regressions on means and variances after each block

	Subject		Adjunct		Whether		CNPC		CNPC w/c	
	M	V	M	V	M	V	M	V	M	V
4 repetitions	.91	.14	.50	.53	.43	.60	.64	.50	.14	.38
6 repetitions	.16	.79	.08	.38	.20	.98	.48	.76	.41	.76
8 repetitions	.31	.60	.72	.81	.17	.75	.84	.32	.26	.35
10 repetitions	.07	.67	.14	.64	.14	.66	.44	.18	.22	.25
12 repetitions	.06	.54	.66	.80	--	--	--	--	--	--
14 repetitions	.14	.83	.52	.26	--	--	--	--	--	--

As Table 4 indicates, none of the regressions reached significance. Adjunct Islands and Subject Islands did trend in the right direction after 6 repetitions and 10 repetitions respectively, but never reached significance. The lack of significant results indicates that there's no evidence of means or variances increasing over time, and thus no evidence of satiation with unlimited responses and balanced designs.⁸

4. Limited response tasks and re-balanced designs

To test the effect of re-balancing the design in a limited response task such as the yes-no task, two experiments were designed using the same superset of materials. The first experiment was modeled after the experiment in Snyder 2000: eight violation types were presented in each of five blocks along with two grammatical fillers. In essence, this is an attempt to replicate the Snyder 2000 results with novel materials (with one additional ungrammatical item per block), and as such the results were reported in Table 2. The second experiment uses the same materials, but only presents four violations per block and six grammatical fillers. Given a definition of satiation that is based on the number of *yes* responses, the equalization response strategy would predict no effect under this re-balanced design: the equalization that occurs would be an increase in *no* responses, and therefore not look like a satiation effect. If satiation is not due to a response strategy, we would predict satiation effects in both the 8 violation design (experiment 1) and the 4 violation design (experiment 2).

4.1 Design

The design was identical to that of Snyder 2000. Participants were presented with 5 blocks of 10 sentences, each block containing either 4 or 8 violations. Sentences were matched for length in number of words within each violation type (but not across violations). The task was a yes-no task.

Table 5: Violations tested (first four are in both experiments)

Adjunct Island	What does Jeff do the housework because Cindy injured?
Relative Clause Island	What did Sue meet the mechanic who fixed quickly?
Complex NP Constraint	What did you doubt the claim that Jesse invented?
Whether Island	What do you wonder whether Sharon spilled by accident?

Coordinate Structure	
Constraint	What did Clare claim she wrote the article and ?
Sentential Subject – Inf	What will to admit in public be easier someday?
Sentential Subject – Fin	What does that you bought anger the other students?
Left Branch Constraint	How much did Mary say that you earned money?

4.2 Results

Responses were analyzed according to the Snyder 2000 definition of satiation: the number of participants whose judgments changed from No to Yes was compared to the number whose judgments changed from Yes to No by the Sign Test:

Table 6: Satiation results for both experiments

	Experiment 1			Experiment 2		
	No to Yes	Yes to No	<i>p</i> -value	No to Yes	Yes to No	<i>p</i> -value
Adjunct	4	1	.38	0	0	1.0
RC	3	4	1.0	2	0	.50
CNPC	1	1	1.0	0	0	1.0
Whether	2	3	1.0	0	2	.50
CSC	3	2	1.0	--	--	--
SS-INF	3	0	.25	--	--	--
SS-FIN	3	1	.63	--	--	--
LBC	4	4	1.0	--	--	--

As the results indicate, and as the discussion in section 2 foreshadows, there were no significant satiation effects *in either experiment*.⁹ These results are not very informative from the point of view of testing the effect of re-balanced designs on satiation as it is defined in Snyder 2000. However, there are other ways to define satiation, or to put it more precisely, there are other ways to define changes in judgments. For instance, a casual glance at the results suggests that there were more data points in experiment 1 than experiment 2, where ‘data points’ are participants with unstable judgments. This suggests

that more participants had unstable judgments in experiment 1 than experiment 2. Fisher's exact test (Fisher 1922) confirms that the difference is significant: in experiment 1, 15 out of 25 subjects show a change in judgment, versus 3 out of 19 subjects in experiment 2, $p < .04$.¹⁰ So while the replication problem prevents a comparison of balanced and re-balanced designs using the definition from Snyder 2000, a more generic definition of judgment instability still shows a significant effect of re-balanced designs in limited response tasks on the stability of participants' judgments.

While it is tempting to simply interpret the stability of the re-balanced design as evidence for the response strategy, the fact that the replication forced a redefinition of satiation adds a layer of complexity. As formulated in section 2, the equalization response strategy is bi-directional: an asymmetry in either direction could lead to the application of the strategy. If this were true, we might actually predict more changes in judgment in the re-balanced experiment than actually occurred, as it is still mildly unbalanced (4 ungrammatical and 6 grammatical items per block). The fact that the re-balanced design was more stable than the canonically unbalanced design suggests that the equalization response strategy may be unidirectional: participants only try to equalize responses when *yes* responses are underrepresented, not when they are overrepresented. While testing this possibility would require additional experiments beyond the scope of the current discussion, it should be noted that the possibility of an asymmetrical equalization strategy is not beyond reason. Naïve participants have most likely never encountered intentionally ungrammatical sentences before, therefore it would not be surprising if ungrammatical sentences were the marked case. A preponderance of ungrammatical items might therefore trigger the type of second-guessing necessary to invoke a response strategy, whereas a preponderance of the unmarked grammatical items might be expected. Regardless of the ultimate status of the directionality of the response strategy, the fact that there is a statistically robust difference in the stability of the judgments across these two designs strongly suggests that the design may be responsible for the instability in Snyder 2000 that underlies the satiation effect, which is predicted by the response strategy explanation, and completely unexpected if satiation is a real property of judgments.

5. Conclusion

This paper has argued that the study in Snyder 2000 included some design elements that may have inadvertently enabled a response strategy, the equalization strategy, that in turn lead to the satiation effect reported there. In an effort to demonstrate the plausibility of this explanation, this paper identified a series of predictions that differentiate between the equalization strategy and a true satiation effect. We have seen that the equalization strategy predicts the replication problem in section 2, the lack of satiation effects with continuous response tasks and balanced designs in section 3, the lack of satiation effects with limited response tasks and balanced designs in section 4, and the correlations between relative acceptability and satiating violations in section 5.

One potential criticism of this conclusion is that it is predicated upon *not* finding an effect, or in other words, on null results. Null results pose a major problem for falsification-based experimentation: lack of evidence for an effect is not evidence for the lack of the effect. In this case, the lack of satiation effects in these experiments does not

necessarily indicate that satiation does not exist. It could be the case that satiation exists, but it is highly variable in the population, and the samples chosen in these experiments happened to contain participants who do not exhibit satiation. While plausible, the distribution of effects suggests otherwise: 3 of the 6 unbalanced yes-no experiments exhibited at least one satiation effect, but 0 of the 6 re-designed experiments demonstrated satiation effects. Furthermore, section 4 presented evidence that the asymmetry between the original experimental design and the re-designs persisted even after redefining satiation to include any change in judgment.

One of the motivations for the original Snyder 2000 study was to determine whether the anecdotal reports of judgments changing over time was in fact a liability for syntactic theory, or whether it was a systematic fact that could be used to refine analyses. Both the equalization strategy explanation and the satiation-is-rare explanation suggest that satiation effects are not an inherent property of violations, therefore satiation effects cannot be a new source of data for syntactic. And while this seems to bolster the original concern that satiation effects undermine the stability of acceptability data, both explanations suggest that satiation effects are not a property of the language faculty proper, but rather a property of individuals. So while it is the case that Snyder's original response to these concerns cannot be maintained, it is also the case that the concerns themselves are unjustified. The experiments in this paper suggest that acceptability judgments are a remarkably stable form of data, even after as many as 14 repetitions.

The study presented in this paper also has implications for the evolving discussion of linguistic methodology that has been unfolding rapidly over the past 10 years or so (Bard et al 1996, Cowart 1997, Featherston 2005, Keller 2000, 2003, Schütze 1996, etc.). For one, it underscores the importance of eliminating the possibility of task-related effects when adopting formal experimental methodologies for acceptability data collection. As we have seen, small decisions such as the choice of response scale and the balance of grammatical and ungrammatical items may affect the results. This study also suggests the need for investigating extra-linguistic factors that may affect the responses of naïve participants, such as the correctability of violations. It is true that professional linguists generally control for factors like these when giving judgments, however, non-linguists may not. Finally, these results may also indicate that unlimited response tasks such as magnitude estimation are more robust to design imbalances than limited response tasks. At the very least, these experiments demonstrate that we do not yet completely understand all of the factors that affect acceptability judgments.

Footnotes

I am indebted to Howard Lasnik, Norbert Hornstein, and Colin Phillips for invaluable comments on earlier drafts, as well as audiences at the University of California San Diego and GLOW 30 at the University of Tromsø. I'd also like to thank two anonymous LI reviewers for many useful comments and suggestions. All errors remain my own.

¹ Hiramatsu 2000 increased the number of repetitions of each violation type from the 5 in Snyder 2000 to 7. Goodall 2005 used novel materials. Sprouse 1 is a direct replication using Snyder's original materials (many thanks to William Snyder). Sprouse 2 uses Snyder's materials but adds a task in which participants also rate their confidence in their judgment. Sprouse 3 uses completely novel materials.

² If the participant shows no change in yes responses, there is no designation. This is because satiation of a violation is defined as comparing the number of satiators to the number of not-satiators. The participants who show *no change in judgment whatsoever* are excluded from the analysis in Snyder 2000 as noise. While this decision has no direct bearing on the replication problem and the equalization strategy, it is worth considering how this limits the interpretation of Snyder's results. By excluding stable participants, Snyder is in essence asking the question: *Of all the unstable participants, were more unstable in a positive direction or a negative direction?* This is a qualitatively different question than *Are judgments stable?*

³ Predictions 2-4 look like a 2x2 factorial design that is missing one cell: unlimited response tasks and unbalanced designs. In fact, those experiments were run as well, with details reported in Sprouse 2007. In order to conserve space, these results have been left out of this discussion. This decision is based on two reasons. First, there is no clear prediction in this cell: the unbalanced design admits the possibility of the equalization strategy, but the unlimited responses make the strategy difficult to implement. Second, the results of the experiments are unclear: if family-wise error is left uncorrected, there are some significant satiation effects, but the effects didn't replicate in a second experiment. This cell raises a set of questions that are beyond the scope of this paper (and orthogonal to the investigation of the equalization strategy).

⁴ There are other benefits, and a few shortcomings, to magnitude estimation. See Bard et al 1996, Keller 2000, and Sprouse 2007a for detailed discussions.

⁵ For a more detailed discussion of the design see Sprouse 2007a.

⁶ The log transformation in linguistic magnitude estimation is intended correct for the rightward skew inherent in responses using the positive number line (Bard et al. 1996, Keller 2000). Sprouse 2007a presents evidence that this may be unnecessary. As such, the analyses were also run on the untransformed data. There were no significant results with the untransformed data either. This isn't entirely surprising: linear regressions are identical to ANOVAs, and ANOVAs are F-tests. F-tests are robust to divergences from

normality as long as the divergence is uniform across all of the conditions. In this case, the log transformation is applied to all of the conditions uniformly.

⁷ The values used for the variance regression were the absolute value of the difference between the responses and the mean, similar to the residuals as calculated in Levene's (1960) test.

⁸ Table 3 indicates that the means and variances increased between some blocks, and decreased between others. In other words the lines of best fit were negative in some cases. Because none of the regressions were significant, the direction of the slope is not represented in the table

⁹ In fact, none of these results were even close to reaching significance. The exact number of responses necessary for significant in the Sign Test varies based on the number of the other responses. So for 0 responses of Y to N, the Sign Test requires 6 responses of N to Y. Other significant pairings would be 8 vs. 1, 10 vs. 2, 12 vs. 3, 13 vs. 4, and 15 vs. 5.

¹⁰ One anonymous LI reviewer suggests that a count of the absolute number of subjects demonstrating a change in judgment may skew the results toward finding an effect in experiment 1 and not experiment 2 because subjects in experiment 1 had twice as many opportunities to change their judgment. While this very plausible a priori, it rests on the assumption that subjects' responses are inconsistent. If anything, the experiments in this Remark suggest otherwise. For example, the magnitude estimation experiments in section 3 provide overwhelming evidence for consistency in judgments over time even when subjects are presented with 14 "opportunities to change their judgment" and a potentially infinite response scale. Given these results, I find it unlikely that 8 repetitions on a binary response scale would be enough to induce inconsistency independently of the ratio of unacceptable to acceptable items. However, for those still unconvinced of the stability of acceptability judgments, it may be worth testing this possibility in the future.

References

- Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72:32–68.
- Braze, Forrest David. 2002. Grammaticality, acceptability, and sentence processing: A psycholinguistic study. Doctoral Dissertation, University of Connecticut.
- Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3), 355-387.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Sage Publications.
- Crain, Stephen, and Janet Fodor. 1987. Sentence matching and overgeneration. *Cognition* 26:123–169.
- Edelman, Shimon, and Morten Christiansen. 2003. How seriously should we take Minimalist Syntax? *Trends in Cognitive Science* 7(2): 60-61.
- Featherston, Sam. 2005. Magnitude estimation and what it can do for your syntax. *Lingua* 115.
- Fisher, Ronald. 1922. On the interpretation of Chi-squared from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* 85:87–94.
- Goodall, Grant. 2005. Satiating and inversion in wh-questions. Talk given at University of Hawaii.
- Hiramatsu, Kazuko. 2000. Accessing linguistic competence: Evidence from children's and adults' acceptability judgments. Doctoral Dissertation, University of Connecticut.
- Kaan, Edith, and Laurie Stowe. ms. Developing an experiment.
- Kaschak, Michael P. & Arthur M. Glenberg. 2004. The construction needs learned. *Journal of Experimental Psychology: General* 133(3), 450-467.
- Keller, Frank. 2000. Gradiance in grammar: Experimental and computational aspects of degrees of grammaticality. Edinburgh: University of Edinburgh dissertation.
- Keller, Frank. 2003. A psychophysical law for linguistic judgments. In Richard Alterman & David Kirsh (eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, 652-657.
- Keller, Frank, M. Corley, S. Corley, L. Konieczny, and A. Todirascu. 1998. Webexp. Technical report hrc/tr-99, Human Communication Research Centre, University of Edinburgh.
- Levene, H. 1960. Robust tests for equality of variances. In *Contributions to probability and statistics*, ed. I. Olkin, 278–292. Stanford University Press.
- Lorch, R. F., Jr., and J. L. Myers. 1990. Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:149–157.
- Luka, Barbara J. & Lawrence W. Barsalou. 2005. Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52, 436-459.
- Noppeney, Uta. & C. J. Price. 2004. An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience* 16(4), 702-713.
- Pickering, Martin J. & Holly P. Branigan. 1998. The representation of verbs: Evidence

- from syntactic priming in language production. *Journal of Memory and Language* 39(4), 633-651.
- Schütze, Carson. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. The University of Chicago Press.
- Snyder, William. 2000. An experimental investigation of syntactic satiation e *Linguistic Inquiry* 31:575–582.
- Sprouse, Jon. 2007a. A program for experimental syntax. Doctoral Dissertation, University of Maryland.
- Sprouse, Jon. 2007b. Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax. *Biolinguistics* 1: 118-129.
- Stevens, Stanley Smith. 1957. On the psychophysical law. *Psychological Review* 64:153–181.

Jon Sprouse
University of California, Irvine
Department of Cognitive Sciences
3151 Social Science Plaza A
Irvine, CA 92606
USA
jsprouse@uci.edu