

## Revisiting Satiation: Evidence for an Equalization Response Strategy

*Jon Sprouse*

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. 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 begin to sound more acceptable after days or weeks of repeatedly judging their acceptability. While linguists have often dismissed this fact as a minor occupational hazard, critics of the use of acceptability judgments have sometimes used it as evidence of the instability of the data underlying syntactic theories. Snyder (2000) replies directly to such criticism with the following logic: If instability is an inherent property of judgments, and not just a symptom of "doing linguistics," then nonlinguists should exhibit it as well. Furthermore, general instability in judgments should manifest itself in every type of unacceptable sentence; that is, 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 exhibited satiation for three out of seven violation types tested. This provided evidence for his hypothesis, and raised the possibility of studying satiation as part of studying 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 nonsatiating violations have entirely different sources within the language faculty. In particular, he suggests that satiating

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 would also like to thank two anonymous *LI* reviewers for many useful comments and suggestions. All errors remain my own.

violations may be due to an extragrammatical factor, such as some sort of processing constraint or resource limitation—a suggestion that receives some support from 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 itself as a higher probability of using that structure in speech (e.g., Bock 1986, Pickering and Branigan 1998) or as faster reading times (e.g., Kaschak and Glenberg 2004, Noppeney and Price 2004). Luka and Barsalou found that syntactic priming can even manifest itself 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 violations of other constraints on acceptability, such as ease of processing (see Sprouse 2007a for discussion).

As this brief introduction indicates, satiation is a topic that touches on 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 article, I reexamine satiation in light of the replication problem, suggesting instead that the satiation effects reported in Snyder 2000 may derive from a response strategy in which participants attempt to equalize the number of times they give each possible response. I argue that this strategy is enabled by two design features of Snyder's original experiment: (a) the task offered only two response choices (*yes* and *no*), and (b) the design included significantly more unacceptable sentences than acceptable sentences. I present nine experiments designed 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 (e.g., Edelman and Christiansen 2003) have suggested.

## 2 The Replication Problem and the Equalization Strategy

### 2.1 Defining the Replication Problem

Snyder (2000) tested the following seven violation types for a satiation effect (see table 1): adjunct island, complex-NP island, left-branch violation, subject island, *that*-trace violation, *want-for* construction, *whether*-island. He found significant satiation for complex-NP islands and *whether*-islands, and marginal satiation ( $p < .07$ ) for subject islands. To the best of my knowledge, since

**Table 1**

Violations tested in Snyder 2000

Violation	Example sentence
Adjunct	Who did John talk with Mary after seeing?
Complex-NP	Who does Mary believe the claim that John likes?
Left-branch	How many did John buy books?
Subject	What does John know that a bottle of fell on the floor?
<i>That-trace</i>	Who does Mary think that likes John?
<i>Want-for</i>	Who does John want for Mary to meet?
<i>Whether</i>	Who does John wonder whether Mary likes?

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.<sup>1</sup> Postponing for a moment a detailed discussion of Snyder's design and definition of satiation effect, I summarize the results of these studies in table 2 to visually demonstrate the replication problem.

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. Replicability is thus crucial to the argument: if some

**Table 2**

Summary of satiation results and the replication problem (✓ = significant effect, (✓) = marginal effect, — = not tested)

	Snyder 2000	Hiramatsu 2000	Goodall 2005	Sprouse 1	Sprouse 2	Sprouse 3
Adjunct						
Complex-NP	✓		✓			
Left-branch						
Subject	(✓)	✓				
<i>That-trace</i>		✓				—
<i>Want-for</i>		✓	—			—
<i>Whether</i>	✓	✓	—			

<sup>1</sup> Hiramatsu (2000) increased the number of repetitions of each violation type from the five used in Snyder 2000 to seven. Goodall (2005) used novel materials. Sprouse 1 is a direct replication using Snyder's original materials (many thanks to William Snyder). In Sprouse 2, I used Snyder's materials but added a task in which participants also rated their confidence in their judgment. In Sprouse 3, I used completely novel materials.

violations satiate in some experiments, and others in other experiments, the effect stops looking like a property of violations and starts looking more like a general property of the task. In the following section, that is exactly what I propose: that 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 reported 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 five blocks of 10 sentences; each block contained one of each of the 7 violations in table 1, and 3 completely grammatical filler sentences. In other words, participants saw 35 unacceptable sentences and 15 acceptable sentences. The operational definition of satiation that Snyder adopted 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* responses increased, the participant satiated on that violation; if the number of *yes* responses decreased, the participant did not satiate on that violation.<sup>2</sup>

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 participants' decision to alter their responses on the

<sup>2</sup> There is no term to refer to participants who showed no change in *yes* responses. This is because satiation of a violation is defined as comparing the number of satiators to the number of nonsatiators. The participants who showed *no change in judgment whatsoever* were 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 it 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 qualitatively different from the question, *Are judgments stable?*

basis of 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. While the equalization response strategy explanation and the satiation explanation make the same predictions for the experimental design in Snyder 2000, their predictions can be teased apart by manipulating the experimental task and the balance of the design.

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. Failures to replicate cast doubt on the view that satiation is a direct reflex of the language faculty. Therefore, the replication problem is evidence in favor of the equalization strategy.

2. *Limited-response tasks and unbalanced designs*

The equalization strategy 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. As noted earlier, judgments are indeed unstable in some experiments of this type. However, the satiation explanation also predicts this.

3. *Unlimited-response tasks and balanced designs*

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

4. *Limited-response tasks and rebalanced designs*

By "rebalancing 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.<sup>3</sup> As we will see, the results favor the equalization strategy.

<sup>3</sup> Predictions 2–4 look like a  $2 \times 2$  factorial design that is missing one cell: unlimited-response tasks and unbalanced designs. In fact, experiments of that type were conducted as well (for details, see Sproue 2007b). To conserve space, I do not discuss the results of those experiments here. There are two reasons for this. First, this cell leads to no clear prediction: 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 did not replicate in a second experiment. This cell raises questions that are beyond the scope of this article (and orthogonal to investigating the equalization strategy).

### 3 Unlimited-Response Tasks and Balanced Designs

Magnitude estimation is an ideal task for giving participants a theoretically unlimited number of potential responses (Stevens 1957, Bard, Robertson, and Sorace 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. For example, as in figure 1, the acceptability of a reference sentence might be assigned an integer rating of 100. Participants would then be asked to rate the acceptability of the target item using the reference as a unit of measure. A participant who felt the target sentence was only half as acceptable as the reference sentence would assign it a rating of 50. A participant who felt it was twice as acceptable would rate it 200. Because the positive number line is infinite, participants theoretically have an unlimited number of potential responses.<sup>4</sup>

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 complex-NP islands. These violations were chosen because subject, *whether*-, and complex-NP islands had shown satiation effects in Snyder's experiment, and because adjunct islands have been analyzed as theoretically similar to subject islands (e.g., Huang 1982). The fifth experiment tested complex-NP 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 five experiments were balanced according to the best practices of psycholinguistics (e.g., Kaan and Stowe 2001): the ratio of acceptable sentences to unacceptable sentences was 1:1, and

<p><i>Reference sentence</i>          What do you wonder whether Mary bought?          Acceptability: 100</p> <p><i>Target sentence</i>          What did Lisa meet the man that bought?          Acceptability: ____</p>
---

**Figure 1**

An example of linguistic magnitude estimation

<sup>4</sup> Magnitude estimation has other benefits and a few shortcomings. For detailed discussions, see Bard, Robertson, and Sorace 1996, Keller 2000, and Sprouse 2007b.

the ratio of fillers to target items was 2:1. To achieve this, the experiments employed a block design in which each block contained 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 three grammatical fillers. The subject and adjunct island experiments presented participants with seven blocks, for a total of 14 repetitions of the island violations. The other experiments presented participants with five 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 (subject islands), 24 (adjunct islands), 20 (*whether*-islands), 17 (complex-NP islands), and 20 (complex-NP islands with context) (compared with 22 in Snyder's (2000) experiment).<sup>5</sup>

### 3.2 Results

The results of each experiment were divided by the value of the reference sentence and then log-transformed following the standard practices of linguistic magnitude estimation (Bard, Robertson, and Sorace 1996, Keller 2000) prior to analysis.<sup>6</sup> Table 3 presents the mean judgment and standard deviation for each block in each experiment.

**Table 3**

Mean judgments (M) and standard deviations (SD) for each block in each experiment

	Subject		Adjunct		<i>Whether</i>		Complex-NP		Complex-NP with context	
	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)						

<sup>5</sup> For a more detailed discussion of the design, see Sprouse 2007b.

<sup>6</sup> The log transformation in linguistic magnitude estimation is intended to correct for the rightward skew inherent in responses using the positive number line (Bard, Robertson, and Sorace 1996, Keller 2000). Sprouse 2007b presents evidence that this may be unnecessary. Therefore, the analyses were also performed on the untransformed data. No significant results emerged with the untransformed data either. This is not 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.

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 task.

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 performed on both the mean judgments and the variances.<sup>7</sup> And since the increases could become 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). As table 4 indicates, none of the regressions reached significance. The results for 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 is no evidence of

**Table 4**

Repeated-measures linear regressions on means (M) and variances (V) after each block

	Subject		Adjunct		<i>Whether</i>		Complex-NP		Complex-NP with context	
	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	—	—	—	—	—	—

<sup>7</sup> The values used for the variance regression were the absolute values of the difference between the responses and the mean, similar to the residuals as calculated in Levene's test.

means or variances increasing over time, and thus no evidence of satiation with unlimited-response tasks and balanced designs.<sup>8</sup>

#### 4 Limited-Response Tasks and Rebalanced Designs

To test the effect of rebalancing the design in a limited-response task such as the yes/no task, two experiments were designed using the same superset of materials. Experiment 1 was modeled after the experiment reported in Snyder 2000: eight violation types were presented in each of five blocks, along with two grammatical fillers. In essence, this was an attempt to replicate Snyder's (2000) results with novel materials (with one additional ungrammatical item per block), and as such the results were reported in table 2. Experiment 2 used the same materials, but presented only four violations per block along with six grammatical fillers. Given a definition of satiation that is based on the number of *yes* responses, the equalization response strategy predicts no effect under this rebalanced design: any equalization that occurs will stem from an increase in *no* responses and will therefore not look like a satiation effect. If satiation is not due to a response strategy, satiation effects are predicted to occur in both the eight-violation design (experiment 1) and the four-violation design (experiment 2).

##### 4.1 Design

The design of experiments 1 and 2 was identical to that of Snyder's (2000) study. Participants were presented with five blocks of 10 sentences, each block containing either 4 or 8 sentences with violations (see table 5). 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 (the first four were tested in both experiment 1 and experiment 2)

Violation	Example sentence
Adjunct	What does Jeff do the housework because Cindy injured?
Relative clause	What did Sue meet the mechanic who fixed quickly?
Complex-NP	What did you doubt the claim that Jesse invented?
<i>Whether</i>	What do you wonder whether Sharon spilled by accident?
Coordinate structure	What did Clare claim she wrote the article and?
Sentential subject (infinitive)	What will to admit in public be easier someday?
Sentential subject (finite)	What does that you bought anger the other students?
Left-branch	How much did Mary say that you earned money?

<sup>8</sup> Table 3 indicates that the means and variances increased between some pairs of 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.

## 4.2 Results

Responses were analyzed according to Snyder's (2000) definition of satiation: using the sign test, the number of participants whose judgments changed from *no* to *yes* was compared with the number whose judgments changed from *yes* to *no*. As the results in table 6 indicate, and as the discussion in section 2 foreshadowed, *neither experiment* showed significant satiation effects.<sup>9</sup> These results are not very informative from the point of view of testing the effect of rebalanced 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, there were more data points in experiment 1 than experiment 2, where “data points” are participants with unstable judgments. This suggests that more participants exhibited 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 participants showed a change in judgment, versus 3 out of 19 participants in experiment 2,  $p < .04$ .<sup>10</sup> So while the replication problem prevents a comparison of balanced and rebalanced designs using the definition from Snyder 2000,

**Table 6**

Satiation results for experiments 1 and 2

	Experiment 1			Experiment 2		
	<i>No</i> to <i>yes</i>	<i>Yes</i> to <i>no</i>	<i>p</i> value	<i>No</i> to <i>yes</i>	<i>Yes</i> to <i>no</i>	<i>p</i> value
Adjunct	4	1	.38	0	0	1.0
Relative clause	3	4	1.0	2	0	.50
Complex-NP	1	1	1.0	0	0	1.0
<i>Whether</i>	2	3	1.0	0	2	.50
Coordinate structure	3	2	1.0	—	—	—
Sentential subject (infinitive)	3	0	.25	—	—	—
Sentential subject (finite)	3	1	.63	—	—	—
Left-branch	4	4	1.0	—	—	—

<sup>9</sup> In fact, none of these results came even close to reaching significance. The exact number of responses necessary for significance in the sign test varies with the number of instances of the other responses. So for 0 responses that change from *yes* to *no*, the sign test requires 6 responses that change from *no* to *yes*. Other significant pairings would be 8 vs. 1, 10 vs. 2, 12 vs. 3, 13 vs. 4, and 15 vs. 5.

<sup>10</sup> One anonymous *LJ* reviewer suggests that a count of the absolute number of participants demonstrating a change in judgment may skew the results toward finding an effect in experiment 1 and not experiment 2 because participants in experiment 1 had twice as many opportunities to change their judgment. While this is very plausible a priori, it rests on the assumption that participants' responses are inconsistent. If anything, the experiments described in this article suggest otherwise. For example, the magnitude estimation experiments reported in section 3 provide overwhelming evidence for consistency in judgments over time even when participants are presented with 14 “opportunities to change their judgment” and a potentially infinite response scale. Given these results, I find it unlikely that eight 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.

a more generic definition of judgment instability still shows a significant effect of rebalanced designs in limited-response tasks on the stability of participants' judgments.

While it is tempting to simply interpret the stability of participants' judgments in the rebalanced 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 bidirectional: an asymmetry in either direction could lead a participant to apply the strategy. If this were true, we might actually predict more changes in judgment in the rebalanced experiment than actually occurred, as it is still mildly unbalanced (four ungrammatical and six grammatical items per block). The fact that participants' judgments were more stable in the rebalanced design than in the canonically unbalanced design suggests that the equalization response strategy may be unidirectional: participants try to equalize responses only when *yes* responses are underrepresented, not when they are overrepresented. While testing the possibility of an asymmetrical equalization strategy would require additional experiments beyond the scope of the current discussion, it should be noted that this possibility is not beyond reason. Since naive participants have most likely never encountered intentionally ungrammatical sentences before, 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 unmarked grammatical items might be expected. Regardless of whether the response strategy turns out to be bi- or unidirectional, the fact that there is a statistically robust difference in the stability of the judgments between 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 yet completely unexpected if satiation is a real property of judgments.

## 5 Conclusion

I have argued that the study presented in Snyder 2000 included design elements that may have inadvertently enabled a response strategy, the equalization strategy, that in turn led to the satiation effect reported there. To demonstrate the plausibility of this explanation, in this article I have identified four predictions that differentiate between the equalization strategy and a true satiation effect: the equalization strategy predicts the replication problem presented in section 2, the lack of satiation effects with unlimited-response tasks and balanced designs presented in section 3, the lack of satiation effects with limited-response tasks and balanced designs presented in section 4, and the correlations between relative acceptability and satiating violations presented in section 5.

One potential criticism of this conclusion is that it is predicated upon *not* finding an effect—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 participants in these experiments happen to be people who do not exhibit satiation. While this explanation is plausible, the distribution of effects suggests otherwise: three of the six unbalanced

yes/no experiments exhibited at least one satiation effect, but none of the six redesigned experiments demonstrated satiation effects. Furthermore, as shown in section 4, the asymmetry between the original experimental design and the redesigned experiments persisted even after satiation was redefined to include any change in judgment.

One of the motivations for Snyder's (2000) original study was to determine whether the anecdotal reports of judgments changing over time posed 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 theorizing. 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 itself, but 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 presented here suggest that acceptability judgments are a remarkably stable form of data, even after as many as 14 repetitions.

The experiments presented here also have implications for the discussion of linguistic methodology that has been unfolding rapidly over the past 10 years or so (e.g., Bard, Robertson, and Sorace 1996, Schütze 1996, Cowart 1997, Keller 2000, 2003, Featherston 2005). 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 extralinguistic factors that may affect the responses of naive participants, such as the correctability of violations. It is true that professional linguists generally control for factors like these when giving judgments; however, nonlinguists may not. Finally, these results may also indicate that unlimited-response tasks such as magnitude estimation are less susceptible 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.

## References

- Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72:32–68.
- Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18:355–387.
- Braze, Forrest David. 2002. Grammaticality, acceptability, and sentence processing: A psycholinguistic study. Doctoral dissertation, University of Connecticut, Storrs.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- Edelman, Shimon, and Morten Christiansen. 2003. How seriously should we take minimalist syntax? *Trends in Cognitive Science* 7:60–61.
- Featherston, Sam. 2005. Magnitude estimation and what it can do for your syntax. *Lingua* 115:1525–1550.
- 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. Satiation and inversion in *wh*-questions. Paper presented at the University of Hawaii.
- Hiramatsu, Kazuko. 2000. Accessing linguistic competence: Evidence from children's and adults' acceptability judgments. Doctoral dissertation, University of Connecticut, Storrs.
- Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, MIT, Cambridge, MA.
- Kaan, Edith, and Laurie Stowe. 2001. Developing an experiment. Ms., University of Groningen.
- Kaschak, Michael P., and Arthur M. Glenberg. 2004. This construction needs learned. *Journal of Experimental Psychology: General* 133:450–467.
- Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Doctoral dissertation, University of Edinburgh.
- Keller, Frank. 2003. A psychophysical law for linguistic judgments. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, ed. by Richard Alterman and David Kirsh, 652–657. Boston: Cognitive Science Society.
- Lorch, Robert F., Jr., and Jerome 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., and 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, and Cathy J. Price. 2004. An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience* 16:702–713.
- Pickering, Martin J., and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39:633–651.
- Schütze, Carson. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31: 575–582.
- Sprouse, Jon. 2007a. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1:118–129.
- Sprouse, Jon. 2007b. A program for experimental syntax. Doctoral dissertation, University of Maryland, College Park.
- Stevens, Stanley Smith. 1957. On the psychophysical law. *Psychological Review* 64:153–181.

*University of California, Irvine*  
*Department of Cognitive Sciences*  
*3151 Social Science Plaza A*  
*Irvine, CA 92606*

*jsprouse@uci.edu*