Acceptability of Comparative Illusions as a function of interactions between Repeatability of a Verb Phrase and Active production task

Long Pham

Faculty of Humanities, University of Amsterdam

Bachelor's Thesis Linguistics

Dr. Titia Benders

June 27, 2022

Acknowledgment

I would like to express my greatest gratitude to my thesis supervisor, Dr. Titia Benders. If not for her guidance and suggestions during the sessions, this paper would have been much less focused and all over the place, and I would not have had the opportunity to develop my topic to this extent. Furthermore, this paper could not have been possible without the Ethics Committee, which approved this study in the first place.

I am also deeply grateful for the efforts of Dirk Jan Vet in creating my experiment from my haphazard notes and lists. Without his expertise in Experiment Designer, this experiment could not have run so smoothly and professionally, and would have been much more difficult for the participants of the experiment.

I would like to recognize my gratitude to all the participants of this experiment, as it would not have been possible without their agreement to participate. In this same vein, I would like to thank my friend Ngan, my aunt Tam, and my colleague Lieke, who went out of their way to help me find additional recruits to meet my initial participant number criteria. Lastly, I would like to thank my colleagues Lieke, Marli, Ceyda, and Issy for their presence during the weekly meet-ups and help alleviate those otherwise potentially lonely writing days.

Table of Contents

Abstract

This study tested the effects of an active production task with the phenomenon of Comparative Illusions which contains verbs that depict repeatable events. Previous research has found an effect of repeatable Verb Phrases increasing the Comparative Illusions' ability to make participants perceive an otherwise grammatically unacceptable sentence as acceptable. However, not many studies have considered the effect of short-term recalling in conjunction with the aforementioned repeatable Verb Phrases on CIs acceptability. This study aims to bridge this gap somewhat by investigating how an interaction of these factors would result in terms of acceptability. It was done by introducing participants to a Comparative Illusion stimulus that either contained or did not contain a Repeatable Verb Phrase, followed by either an active production task via re-typing or a probe question. The experiment yielded significant results on the effects of Repeatable Verb phrases. However, no significant results were found on the effect of active production tasks, or their interactions in the acceptability of Comparative Illusions. Therefore, whether or not actively producing Comparative Illusions that contain repeatable Verb Phrases would affect its acceptability, and by extension, the strength of the Comparative Illusion, was inconclusive.

1. Introduction

Consider the sentence in (1). Native English speakers have given (1) generally high scores in terms of how acceptable it is as an English sentence (Leivada, 2021; Wellwood et al., 2018). When tasked with generating meaning for the sentence, however, speakers often become confused (Leivada 2021).

(1) More people have been to Russia than I have.

This sentence type is called a Comparative Illusion (abbreviated as CI). As the name implies, it is, first and foremost, a comparative structure – it compares two clauses with each other. Most research addresses CIs as sentences where at first the sentence was considered acceptable by the standards of their language. However, upon closer reflection, Comparative Illusions have no stable, meaningful interpretation (Wellwood et al., 2018; Leivada, 2021). In other words, they only have a superficial appearance of well-formedness. In (1), readers can infer from "more people" that they are likely to require a certain interpretation of an operator – an expression that denotes syntactic movement - of a variable in a dependent than-clause. For example, if "than I" in (1) was replaced by "than elephants", the variable can be posited in the determiner position of the bare plural, and the rest of the sentence can be parsed normally. However, the phrase "than I" like in (1) does not allow for the completion of the dependency, therefore the parser must wait for a suitable nominal correspondent. If the sentence continued with the phrase "than I expected", for example, the gap after that could be seen as part of the elided clause, but this was not possible in (1) (Wellwood et al., 2018). It is thus striking that CIs seem to be as acceptable as informally reported.

While the appearance of well-formedness is an effect not unique to Comparative Illusions, what separates that sentence type from other forms of illusions, such as negative polarity items illusions, is that Comparative Illusions never arrive at a specific, grammatically-licensed interpretation, as there does not seem to be a single misinterpretation that speakers converge upon. Instead, they continually struggle to articulate the interpretation while believing it to have a coherent interpretation. Whereas in (2), which is an NPI illusion (Orth, Yoshida & Slogett, 2021), an interpretation (albeit incorrect) could be reached in (3), it is not possible for (1) to reach the same kind of conclusion.

(2) The man who no woman trusted will ever go to Paris.

(3) The man who no woman trusted will go to Paris.

The inability to arrive at a single interpretation that Comparative Illusions have suggested the involvement of online processing within the phenomenon, Wellwood et al. further stated. Therefore, the current research aimed to help shed some further light on this phenomenon by investigating the different factors regarding the acceptability of Comparative Illusions and how these factors interact with each other, especially within the realms of sentence processing through the involvement of short-term recall.

1.1. Theoretical background

Previous experiments in general focused on testing the robustness or the existence of the Comparative Illusion effects – where it was found to make participants judge a comparative structure sentence as grammatically acceptable while it was not. It has been shown that the effect seemed to be robust (Philips, 2011; Wellwood et al., 2018). The existence of this effect is mostly manifested in acceptability judgment tasks.

1.1.1 Factors leading to the effectiveness of Comparative Illusions

There have been studies about the factors that affect how acceptable Comparative Illusions are. Christensen (2016)'s study included grammatical number of the *than*-phrase subject. Similarly, Wellwood et al. (2018)'s experiments included a variety of factors such as comparative quantifier within the main clause subject position, whether VP ellipsis was applied in the *than*-clause, and whether or not the denotation of the than-clause subject could be included within the subject NP. However, this current study would focus on the Repeatability of events depicted in the Verb Phrase, which is one factor that had been more consistently present in studies regarding Comparative Illusions. An example of a Repeatable VP was in (1), in which the VP "have been to Russia" is repeatable, as one can be to Russia multiple times. On the other hand, a VP such as "graduating high school" would be not repeatable, as one cannot graduate high school multiple times.

One of the more referenced studies that leaned in favour of Repeatable Verb Phrases as a factor that affects Comparative Illusions was by Wellwood et al. in 2018. This study outlined four possible hypotheses on the causes of Comparative Illusions - Syntactic template matching; Ellipsis repair; *more* ambiguity; and event comparison. Of the four, Wellwood et al. subscribed to the theory of the event comparison. This theory relates to two different readings of a comparison sentence - an individual counting and an event counting. Examples of these two readings through (4) and (5) were first described by Krifka (1990). If we follow the individual counting, we would say that there were 10 individual sailboats that passed through the lock compared to only 5 barges, therefore (5) would be considered correct. However, if the event counting reading is used, then (5) would be considered incorrect, because there were more events from the same barges that passed the locks compared to the sailboats.

(4)   10 sailboats passed through the lock 10 times each (100 passings), and 5 barges passed through the lock 50 times each (250 passings).

(5)   More sailboats passed through the lock than barges did.

Through the link to the event counting interpretation by Krifka (1990) above, Wellwood et al. (2018) then stated the event counting hypothesis, in which the Comparative Illusions reflect speakers' attempts to compare the number of events that occurred within the sentence. This can be related back to (1), where people may fleetingly interpret it as there have been "more people" that have done the event of being in Russia than the "I" pronoun has. This

hypothesis predicted that the effect should be facilitated when the semantic properties of the VP support an event-counting interpretation as opposed to the individual-counting interpretation. In turn, this predicts that the VP must be 'repeatable'.

Wellwood et al. first started with acceptability judgment studies. Their focus was on the manipulations of the Comparative Illusions to test the different theories as to the cause of its effect. Their first experiment tested responses based on a 7-point scale, with 1 being 'unacceptable', and 7 being 'acceptable'. The factors that they tested were the manipulations of the illusion and control conditions (coded as Comparative) crossed with the subset of factors that were borne from the previously mentioned theories. Control conditions were described as sentences with bare plural subjects – plural subjects that lack an overt article or determiner (De Swart & Zwarts, 2009), shown in (6); while the illusion condition were subjects that were not a bare plural, such as non-bare plural subjects shown in (7) or bare singular subjects. Within this subset of factors, the Repeatability of the VP (coded as Repeatability in Wellwood et al.) would be focused on in the current study.

(6)* More girls ate pizza than **boys** did.

(7)* More girls ate pizza than **the boys** did.

The results showed a higher acceptability rating in the control condition than in the illusion condition. Most importantly to the current study, there was a significant interaction only between Repeatability and Comparative. In other words, this experiment found that only the Repeatability of a VP would affect the acceptability within the illusion condition more than within the control condition. Therefore, they concluded that the only hypothesis supported by the results is the event comparison hypothesis, as the hypothesis predicted the acceptability of Comparative Illusions to depend on the possibility of an interpretation of a repeatable event.

However, there has been research that rejects the event-counting hypothesis as well, and subsequently the Repeatability of the Verb Phrase as a factor affecting how acceptable Comparative Illusions are. Christensen's (2016) paper argued the effect of CIs is dependent on the repeatability of the VP and the grammatical number of the *than*-phrase subject. The results showed that within the error rates, where people judged the grammatical illusions as acceptable, only the plurality of the *than*-subject had a significant effect. In other words, having the *than*-subject be a plural (ie. "workers" or "people") would increase the error rate of the acceptability judgments. On the other hand, there was only a non-statistically significant tendency for Comparative Illusions with repeatable VP to be judged as more acceptable.

1.1.2. *Effects of short-term recall on Comparative Illusions*

One angle of approach to the potency of Comparative Illusions is to see how its active production would make a difference. The reason for this is because active production immediately after the presentation of stimuli would mean that there is short-term recalling involved. According to standard views regarding short-term memory, it entails a short

representation of the surface form of information, such as phonological, orthographical or articulatory information (Potter & Lombardy, 1990). However, this standard view of short-term memory of sentences was challenged by Potter & Lombardy (1990). As an alternative, they proposed that immediate short-term recall of a sentence was not based on surface representation but also the meaning of the sentence, similar to long-term sentence recall. The reason behind the verbatim nature of short-term sentence recall would then be because recently activated lexical entries would be the most likely to be selected to express the concepts of the sentence. In other words, the words that were used in a sentence that was immediately presented before would have a high chance of being lexically recalled.

This experiment's results, that being a nearly verbatim recall of the sentences, supported the claims of short-term memory also recalling the conceptual forms. There was a marked increase in the "lure" words intruding in the recall task, thus proving that recent lexical activation makes a word more available for use in immediate recall. Additionally, because there was a tendency for suitable "lures" to intrude even when it was not on the distractor list, they concluded that the short-term recall of the sentence is based on the representation of the sentence meaning.

The results and conclusions of Potter & Lombardy (1990) were built upon by Wellwood et al. (2018), who applied Potter & Lombardy's notions of short-term recalling affecting conceptual forms also with Comparative Illusions. In Experiment 3, they hypothesized that the patterns within the acceptability data that were exhibited in Experiments 1 & 2 (Experiment 1 was elaborated in 1.1.1) can be explained through the possibility of Comparative Illusions to have an event-counting interpretation. Therefore, Comparative Illusion sentences, which support the event-counting reading, were predicted to be easier to recall for participants than for those who do not. This was because recalling the form of the sentence is possible if meaning can be stored for it. Specifically, participants will try to recover the meaning of the target sentence during the recall task, and this will be more difficult if the sentence is harder to interpret or it is less acceptable.

This experiment manipulates the Comparative factor and Repeatability factor. The way that they manipulated the Repeatability factor differs in two sets of items, one of which where the Aspect of the predicate was manipulated. In these items, they were classified as non-repeatable if they had an initiative or terminative aspectual verb to introduce their VP, otherwise repeatable if they had a continuative aspectual verb or a form of "be". They provided examples in (8) and (9), where (8) is an example of non-repeatable VP and (9) is an example of repeatable VP. The other set of items had the Object item modified, however, this is not very relevant to the current study.

(8) Mary **{started, finished}** reading the book.

(9) Mary **{continued, was}** reading the book.

The procedure in this experiment is very similar to Potter & Lombardi's (1990)'s experiment, with some clarifications. There would first be a fixation cross for 500ms, followed by a visual mask. During the Sentence phase, words of a sentence appeared in a rapid serial visual representation (RSVP) mode of 200ms per word, followed by another visual mask. A list of 5 words appeared in RSVP for 250ms per word, followed by a third visual mask. capitalized word would appear for 500ms, and participants were then asked whether or not the word was in the previous list. Finally, they were asked to recall the sentence. This process is visualised in Figure 1.
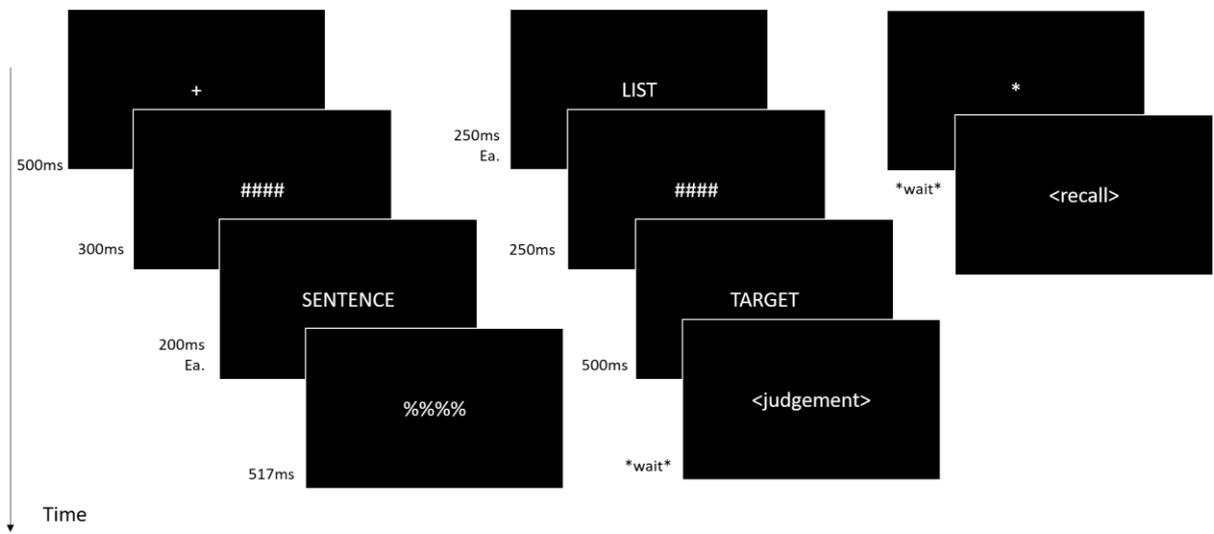

Figure 1: Recreated version of Wellwood et al. (2018)'s Experiment 3 procedure

The results showed a higher failure rate of participants to recall sentences on illusion conditions than on control conditions. However, the illusion stimuli with non-repeatable VPs illusions were not more difficult to recall than illusion stimuli with repeatable VPs. When participants were able to recall the target sentences, they made more changes in the repeatability and number parameter in illusion conditions than in control conditions. This proved a link between the acceptability and the recall data - the less acceptable the Comparative Illusion was; the more modifications were needed in order to have the meaning of the sentence. When counting errors within successful recall trials, there were more errors for sentences with the illusion condition than on the control condition. However, there was no effect of the Repeatable VPs in making errors during recall, and there was no interaction between the factors Comparative and Repeatability. Because participants were found to tend to assign a meaning to even an unacceptable sentence, they concluded that the 'event comparison' reading would be the most relevant predictor of speakers' fixation on the meaning.

If short-term recall was used for Comparative Illusions, it may result in an accurate retrieval of Comparative Illusions. Following Potter & Lombardy's (1990)'s conclusions, both the surface form and the meaning would be retrieved after a short-term recall. Therefore,

participants might realize that the Comparative Illusions' grammaticality was false if they actively produce it. In addition, Wellwood et al. observed more errors made for Comparative Illusions than in control sentences during sentence recall tasks. In turn, this lends support to the idea that active production tasks, which are in essence short-term recall tasks, would affect how Comparative Illusions are interpreted.

To conclude, repeatable Verb Phrases within stimuli as well as the usage of active production tasks, as separate factors, have been studied before in terms of their effect on how Comparative Illusions can be affected in terms of acceptability. However, their interactions and their results on the acceptability of Comparative Illusions have not been studied as far as these articles have been presented. While Wellwood et al. (2018)'s study had included both elements in its experiment, the crucial difference in terms of purpose between that study and my current study was that the recall factor was what they were measuring. This presents a gap in our current knowledge of the phenomenon.

1.2. Hypothesis and predictions

With these articles in mind, this current study attempted to answer: how does self-production affect the impact that repeatable VPs have on Comparative Illusion effects. This experiment answered the research question through a combination of an active production task, where participants must actively produce the Comparative Illusion, as well as afterwards testing their offline well-formedness or acceptable judgement of those sentences. If the event-counting hypothesis as shown in Wellwood et al. (2018) held in this experiment, then it stands to reason that the repeatability of the VP would have an effect on both the repeat task and the comprehension task.

Furthermore, the main hypothesis of this research question is about the interaction between the Repeatability of the VP and the inclusion of the self-production task. I predicted that actively producing stimuli with repeatable VP would make a Comparative Illusion less acceptable compared to not doing so. In other words, actively producing the stimuli with repeatable VP would make the Comparative Illusion effect significantly less potent compared to actively producing non-repeatable VP within the stimuli. Conversely, not actively producing the stimuli would not affect the error rate much, regardless of whether or not repeatable Verb Phrases were included within the stimuli.

2. Methodology

2.1. Design

Following the research question, this experiment had a 2 by 2 factorial design. Both of the independent variables - the repeatability of the VP ([+/-repVP]), which would be elaborated in the Stimuli section; as well as the inclusion of an active production task ([+/-selfprod]), which would be elaborated in the Procedure section - are presented to all participants. Therefore, they are within-subject variables. The dependent variable was the acceptability of the target

stimuli. This was measured via error rates, which would be further elaborated within Data Analysis.

2.2. Participants

Participants were selected to be either native English speakers of any dialect or have at least more than 5 years of experience in English usage. The latter of the two, while was not a criterion that was included in other studies, was used here because it widened the recruitment possibilities. The age of the participants was not a crucial criterion, nor have I seen a definite range in testing the effect in adults in previous papers regarding Comparative Illusions, but to keep a consistent age range between participants, recruited participants' ages ranged from 18 to 40. Snowball sampling was used to recruit participants – participants were first recruited through the connections of the experimenter; whom they know to fit the criteria, then afterward through the connections of said connections. This recruitment method is not perfect, as elaborated further in the Discussion.

These participants were sent an email containing the link to the experiment, as well as their subject number, which determines which version of the experiment they were going to take. Responses from 23 participants were recorded (N=23). Within those responses, one participant encountered an error during the procedure, thus their responses were disregarded. Additionally, 7 more participants (N'=7) agreed to the participant and were sent links and subject numbers to the experiment, however, they did not manage to respond by the end of the data collection period. The total number of participants in the experiment thus was 30.

2.3. Stimulus

The target stimuli, all coming in the form of Comparative Illusions, tested the effect of the Repeatable VP factor. These stimuli were taken from O'Connor (2015) and Wellwood et al. (2018). None of the stimuli from these two sources overlapped. According to Wellwood et al. (2018)'s stimuli design, having the stimuli sentences ranging from 11 to 18 words ensured a suitable level of difficulty without being overwhelming to the participant. As many of O'Connor's (2015) stimuli exceed the 18-word limit, this study reduced the amount by cutting off what was considered "unnecessary", mostly adjunct phrases that do not change the meaning of the main sentence should the phrase be removed. (10) and (11) illustrate this change – whereas the original sentence from O'Connor in (10) might have been too long for the participant, especially within the re-typing task, (11) reduced the number of words by omitting an unnecessary part of the sentence. A total of 30 target stimuli were used during this experiment, divided into two counterbalanced lists. Examples of these target stimuli are shown in Table 1.

(10) This year, more convenience store clerks bought jackpot lottery tickets than my customer did because clerks got the tickets at discounted prices.

(11) This year, more convenience store clerks bought jackpot lottery tickets than my customer did.

Table 1: Example target stimuli of 4 target conditions between two lists

List 1			List 2			
TargetSent	SelfProd	RepVP	TargetSent	SelfProd	RepVP	ProbeSent
This semester, more girls ate pizza than boys did.	+	+	This semester, more girls wrote their first haiku than boys did.	+	-	There were girls who ate pizza.
More financial analysts hired from Harvard's business school than our manager did.	-	+	More financial analysts graduated from Harvard's business school than our manager did.	-	-	There were financial analysts hired from Harvard's business school.
Last spring more ducks hatched from eggs than my chicken did.	+	-	Last spring more ducks laid eggs here than chickens did.	+	+	There were ducks who laid eggs here.
More young people finished reading War & Peace this month than old men did.	-	-	More young people resumed reading War & Peace this month than old men did.	-	+	There were young people who resumed reading War & Peace.

In addition, following the stimulus design of Christensen (2016), there were three filler stimuli conditions: Complementizer Phrase comparison (CP-comp), where the sentences were well-formed event comparisons, Determiner Phrase comparison (DP-comp), which was nominal comparisons between subjects, and Ungrammatical Sentences (*Ungram), to which Christensen stated that it contained "word salad permutations of phrases from target conditions". Similar to target stimuli, would be dispersed within the two lists themselves. In total, 10 filler items would be used in each list as well, with examples shown in Table 2.

Table 2: Example filler stimuli of 4 target conditions between two lists

List 1			List 2			
TargetSent	SelfProd	RepVP	TargetSent	SelfProd	RepVP	ProbeSent
As many aunts attended my sister's wedding as uncles did. (DP-comp)	-	+	As many aunts attended my sister's dinners as uncles did. (DP-comp)	-	-	There were aunts that attended the sister's dinners.
Last fall, more engineers relocated to San Francisco than elsewhere. (CP-comp)	+	-	Last fall, more engineers traveled to San Francisco than elsewhere. (CP-comp)	+	+	There were more engineers that traveled.
* More football players were visit the Hall of Fame this year invited to than the baseball player was (*Ungram)	-	+	* More football players were the Hall of Fame this year formally inducted into than the baseball player. (*Ungram)	-	-	There were football players who visited the Hall of Fame.

2.4. Procedure

To reach a wide range of participants, this experiment was conducted through an online medium using the Experiment Designer program. A countdown of 5000ms before the experiment begins first appeared for participants to be ready.

Each trial began by showing the sentences as a whole for a total of 3000ms. The stimuli presentation was followed by a visual mask for 500ms. Next, the elicitation condition was tested. Depending on the indication on the stimuli list associated with a stimulus, participants were given instructions on the screen to either re-type the sentence or to answer an affirmative/negative mono clausal probe question. After this task, an asterisk appeared on the screen for 500ms, and participants were asked to judge whether or not the sentence was well-formed or not. Once the judgment task was finished, the trial was considered to be complete, and the process was repeated with a new stimulus. A schematic of each trial could be found in Figure 2.
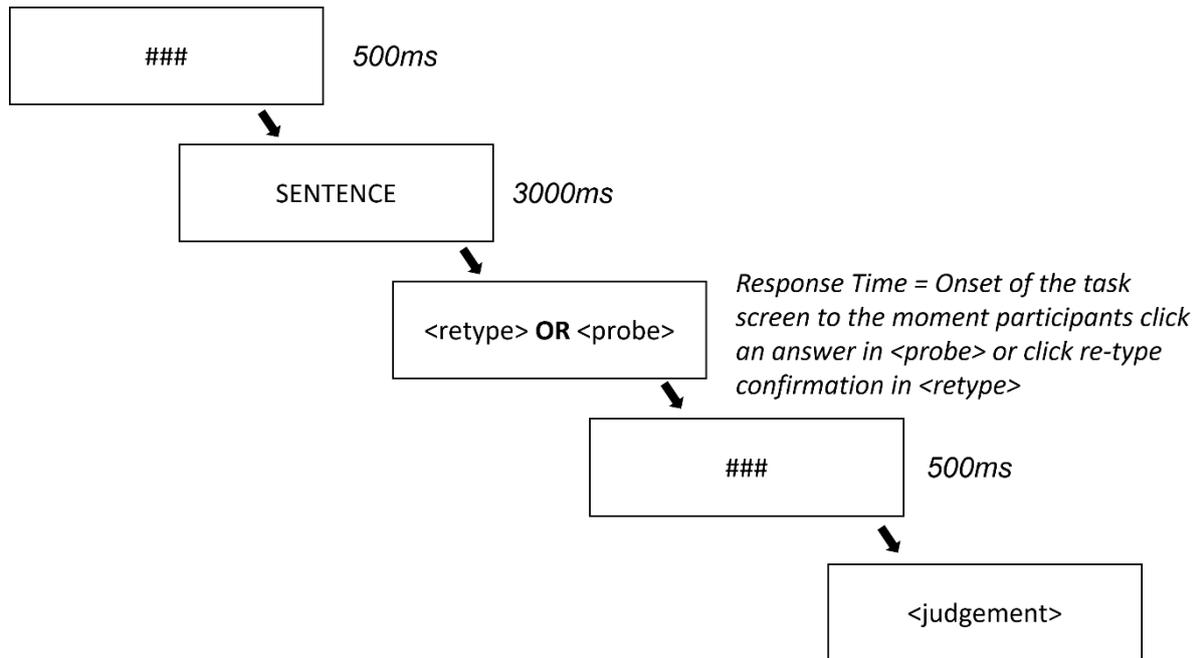
Figure 2: The procedure of a singular trial in the current experiment

In accordance with the previous experiments on acceptability judgment tasks, the response times were recorded for each stimulus from the onset of the task screen until the moment the participant either clicks on the "yes" or "no" answer in the case of the probe question task; or the moment the participant clicks on the confirmation of their re-type in the case of the re-typing task. This was useful as a method of screening the participants' answers, as explained in the screening methods in 2.5.

The self-production factor was determined by the task that the participants performed following the presentation of a stimulus. The task would be determined by the stimulus list. In other words, each stimulus had a task attached to it. The target stimulus would be divided into two lists to ensure the participants only see the variations of a single Comparative Illusion only once. Participants that were given an odd subject number (ie. 1, 3, 5, etc.) would be assigned List 1, while participants with an even subject number (ie. 2, 4, 6, etc.) would be assigned List 2.

Unfortunately, the experimental procedure was not borne out properly. An overlooked error in the coding in Experiment Designer meant that the participants only received List 1 instead of counterbalanced stimuli lists - where subjects with odd numbers receive List 1 and subjects with even numbers receive List 2. The implications for the results of this error would be discussed in section 4.2.

2.5   Data Analysis

As previously mentioned, the measurements that this experiment will be taking were the error rates within the acceptability judgment tasks. As all of the target sentences are grammatically ill-formed, the "correct" answer was to judge each target as unacceptable. Therefore, the error rates, in this case, were the proportion of "acceptable" judgments out of

the total judgments. These data were analyzed separately using an ANOVA model mainly, with the independent variables being [RepVP] and [Selfprod]. However, a generalized linear model was also used to run through the data again. Both of these methods utilized the software R. While participants were potentially tasked with a probe question, their answers did not count towards the results, and thus, they would not be analyzed.

In addition, the response times for the tasks serve as a basic indication of whether or not the participant was truly performing the experiment. If consistently too little time elapsed between the presentation of the stimuli and the answer, for example only a few hundred milliseconds after the onset of the task screen for a probe question, it would be considered a very quick answer. These answers would be considered for removal if it was especially quick compared to the response time of other participants. This also applied to using a single answer for every question, where each question was given only one answer, either with "yes" or "no", and the retyping task was left consistently sloppy. However, no participants were removed via this method.

3.   Results

The hypothesis was that there should be a significant interaction between the type of task and the repeatability of the VP. Actively producing stimuli with repeatable VP should decrease the error rate compared to not doing so. In addition, actively producing non-repeatable VP conditions will make the Comparative Illusion effect more easily seen through compared to stimuli with the probe-question task. Therefore, it was predicted that the [-RepVP].[+Selfprod] category would score a significantly lower error rating compared to the [-RepVP].[-Selfprod] category. In other words, there should be a main effect of the self-production task ([Selfprod]) - the error rates of both the repeatable and non-repeatable Verb Phrase conditions would decrease if they were self-produced compared to if the conditions were not. There should also be a main effect of [RepVP] where having that type of verb would make the acceptability of the Comparative Illusion higher than without. This would correspond to the expected results based on literature by Wellwood et al. (2018).

Because every participant had been assigned the same list, they will all be grouped and analyzed together. Figure 4 shows the error rate in the different combinations of independent variables. The dotted lines in all four categories represent the range of error rates of those categories, while the black bold line in every condition represents the median. The boxes in the four categories represent where the 50% of the error rates lie While the error rate of participants in most categories ranged greatly from 0% to 100%, it can be seen that generally, participants had a higher error rate under [+RepVP] conditions compared to [-RepVP] conditions. This is supported by the median of the four categories, as the median error rates for both [+RepVP] conditions were slightly above 40%. Comparatively, the [-RepVP] condition medians were around 25% for [-RepVP].[-Selfprod] and 18% for [-

RepVP].[+Selfprod]. Furthermore, there does seem to be a pattern of interaction between [Selfprod] and [RepVP], as the error rates in the higher ranges seem to decrease more between [+RepVP].[-Selfprod] and [+RepVP].[+Selfprod] compared to between [-RepVP].[-Selfprod] and [-RepVP].[+Selfprod].
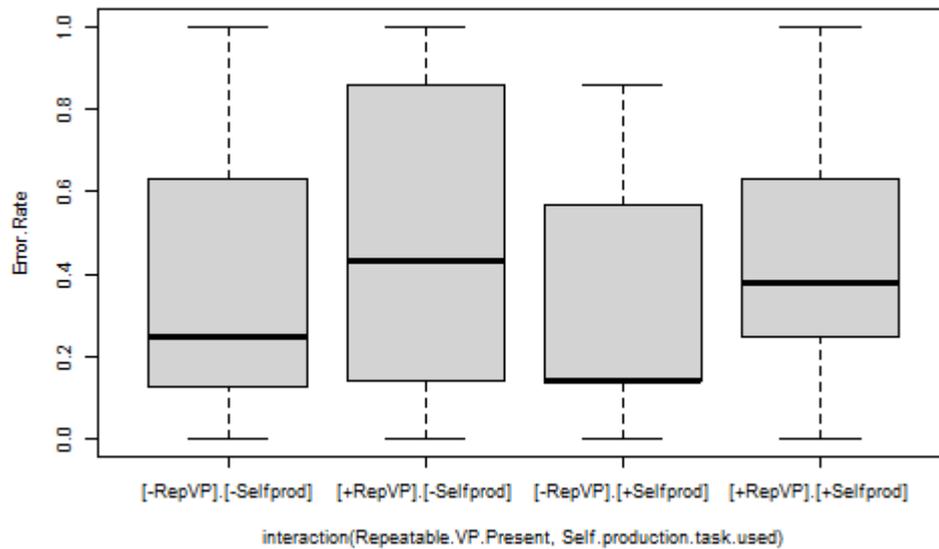

Figure 3: The error rates within each condition of the target stimuli

However, based on the means of each category, as shown in Table 3, the error rates within each category were not entirely as predicted. While the error rates within the [+RepVP].[-Selfprod] condition was the highest (48%) and the [-RepVP].[+Selfprod] was the lowest (33%), the error rate of the [+RepVP].[+Selfprod] was much higher than expected, surpassing slightly the [-RepVP].[-Selfprod] (46% versus 41%). This is in contrast to my predictions of these categories' error rates in relation to each other, with the [+RepVP].[+Selfprod] condition scoring a much lower error rate than [-RepVP].[-Selfprod] condition. These results help prove that there was an effect of Repeatable Verb Phrases in making Comparative Illusions more acceptable.

Table 3: Average Error Rates of Target Stimuli and Filler Stimuli

	Average Target Stimuli Error Rates	Average Filler Stimuli Error Rates
[-RepVP].[-Selfprod]	0.41	0.55
[+RepVP].[-Selfprod]	0.48	0.3
[-RepVP].[+Selfprod]	0.33	0.2
[+RepVP].[+Selfprod]	0.46	0.38

Statistical analysis did not find any significant effect in either the [Selfprod] $(F(1,80)=0.018, p=0.893>0.05)$; [RepVP] $(F(1,80)=1.755, p=0.189>0.05)$ as independent variables. Most important to the research question, however, their interactions in affecting the error rates of perceiving CIs were not significant $(F(1,80)=0.622, p=0.433>0.05)$. As a double-

check, the analysis was run through the generalized linear model in R, with the Independent variables being the Repeatable Verb Phrases (RepVP) and Self-production task (Selfprod), and the dependent variable being the error rate (ErrorRate). The results were in line with the ANOVA analysis, with none of the independent variables having a significant effect on error rates (RepVP: p=0.707>0.05; Selfprod: p=0.938>0.05); and no effect of interaction between the two variables on error rates as well (RepVP:Selfprod: p= 0.975>0.05). This meant that from these types of analyses, cannot be concluded that there was an effect of Repeatable VP stimuli on the active production of Comparative Illusions.

However, the data that was processed was also not normally distributed. A Shapiro-Wilk test showed that none of the independent variables had an effect on the error rate with the p-value equal to or above 0.05. This meant that the null hypothesis that the data sample comes from a normal distribution (King & Eckersley, 2019) could be rejected. In turn, this meant that the ANOVA and the Generalized Linear Model analyses were not applicable. Therefore, an Aligned Rank Transform was used to analyze the data, as it could perform nonparametric analyses as well as repeated measures (Wobbrock et al., 2011). This was performed in R with the package ARTool. The results showed a non-significant effect of self-production tasks (F(1,63)=0.277, p=0.600>0.05), and the interaction effect between the two independent variables (F(1,63)=2.010, p=0.152). This was in line with the results from the ANOVA analysis. However, a significant effect of Repeatable Verb Phrases in stimuli (F(1,63)=10.001, p=0.002<0.05) was found. This significant effect of Repeatable Verb Phrases within the stimuli was in contrast with the ANOVA analysis and corresponds to the findings of Wellwood et al. (2018).

In addition to the target stimuli being the main target of analysis, participants' results regarding filler stimuli could be examined as well to see if participants can detect the grammaticality better within the few non-Illusion Comparative constructions in contrast to Comparative Illusions. With the filler stimuli, however, the scoring is different – while error rates of the test stimuli depend on the number of 'yes' answers, the error rates of the filler stimuli depend instead on the 'no' answers since most of the filler stimuli are actually grammatical. There was an exception in the case of *Ungram filler condition – since these sentences were deliberately ungrammatical, the error rates within this condition would depend on the 'yes' answers, similar to the test conditions. Therefore, the error rate within this filler condition is determined by subtracting from 1. Because the experiment was not counterbalanced in the end, every *Ungram filler condition also coincides only with the [+RepVP].[-Selfprod] condition. Compared to the testing stimuli, the average error rates of the filler stimuli generally range from slightly higher (particularly the [-RepVP].[-Selfprod]) to moderately lower than the test stimuli. This is shown in figure 4. This tentatively shows that participants are somewhat better at detecting grammaticality within non-Illusion Comparatives.
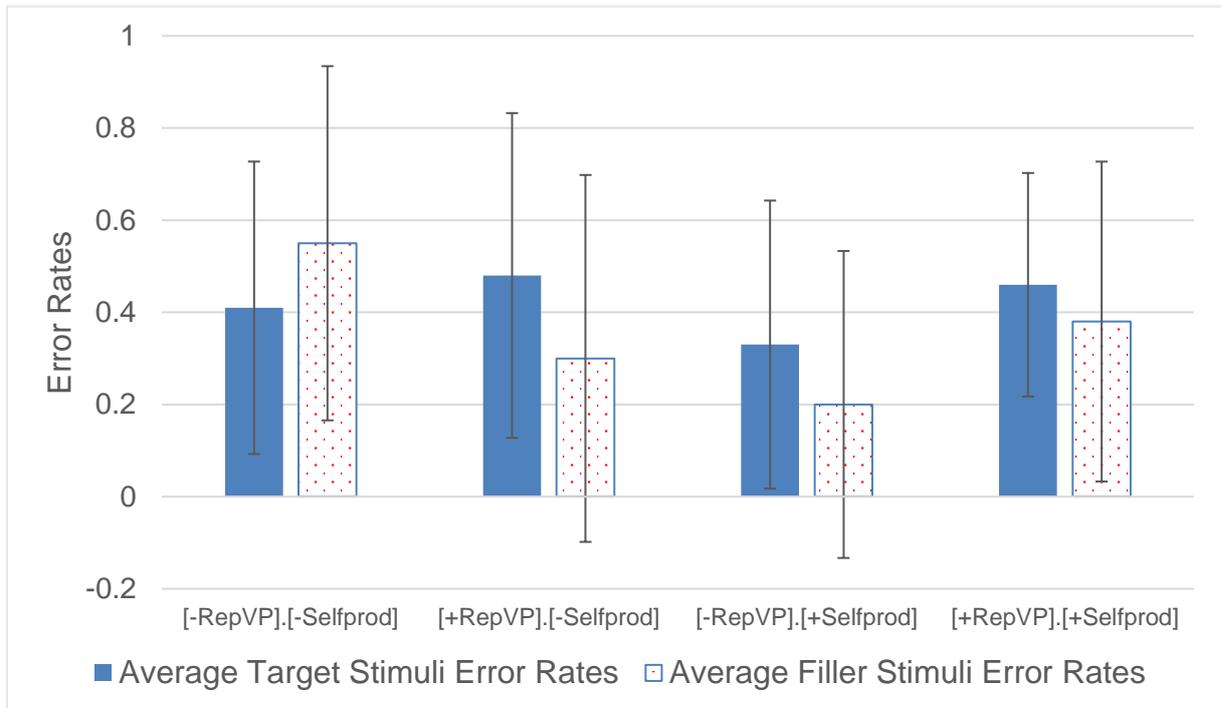
Figure 4: The average error rates between the test stimuli trials and the filler stimuli trials (left blue columns indicate test stimuli trials, right dotted columns indicate filler stimuli trials)

4. Discussion

From the results, participants gave results that were somewhat in line with the previous literature. In particular, a higher error rate in acceptability judgments corresponded to the inclusion of Repeatable Verb Phrases within the test stimuli. However, the prediction that using active production tasks would decrease the error rates in Comparative Illusions was not observed. Furthermore, there was no significant evidence that actively producing stimuli with repeatable VP would make the Comparative Illusion less acceptable compared to not doing so. In this section, I will discuss the differences between the current experiment with those before and the possibility of those affecting the results of the current experiment.

4.1. Differences from previous studies as an explanation of difficulty within trials

A recurring observation from participants was that their tasks were difficult. They claimed in particular that the sentence stimuli were too long for them to remember, and for multiple trials, they did not have enough time to read through them before doing the tasks related to the stimuli. At least one participant claimed that they did not know how to answer many probe questions. While written recalls or probe question responses to the tasks were not hypothesized to affect the acceptability rating, the difficulty of the tasks could make participants wearier going through the experiment. While that also does not result in the effects being invalidated, it generally should be avoided if the experiment does not intend to be

causing fatigue in participants. With this in mind, this section would explore the differences in methodology from the previous experiments leading to the possibility of fatiguing participants.

The decision to have the stimuli ranging between 11 to 17 words in length was based on previous literature by Wellwood et al. (2018). The authors themselves stated that the stimuli should be at that length in order to ensure the difficulty of the recall task in particular. Therefore, it is unlikely that the sentence length causes a problem that affects the acceptability judgments.

This brings to the matter of the differences between the studies in terms of the mode of stimuli presentation, in which there was also deviation. In Wellwood et al.'s (2018) experiment, every single word of the stimuli was presented to the participant for 250ms before the visual mask. This usage of Rapid Serial Visual Presentation was originally from Potter & Lombardy (1990), who stated that 250ms is close to the normal reading rate for college students specifically. Since the length of the stimuli sentences in Wellwood et al. (2018) was from 11 to 17 words long, as previously stated, with an average of 14.2 words, the total time to show the stimuli was 2750ms to 4250ms, with an average of 3550ms. By contrast, the current experiment presented the stimuli in their entirety at the same time, for a total of 3000ms. This meant that participants get even less time to read through the stimuli in the current experiment than even the time needed to read stimuli at the average length within Wellwood et al.'s experiment (2018). Therefore, it makes sense why participants might feel that there was not enough time to read through the stimuli. A possible consequence because of this is that participants would not have enough information to recall in the first place, which would result in guessing.

However, could presenting the sentence in its entirety affect this assessment by the participants as well? As I previously stated, Wellwood et al. (2018) followed the procedure in Potter & Lombardy's (1990) experiment, which also entailed the usage of rapid serial visual presentation. There was no specification as to the reasons Potter & Lombardy used RSVP in their methods. However, their hypothesis was that "verbatim" sentence recalling was based on the conceptual representation of sentence message and partially on unordered recent activation in the lexicon. Therefore, the usage of RSVP makes sense in their experiment because RSVP would perhaps highlight the individual words that would be activated within the lexicon. Following that, Wellwood et al. (2018)'s experiment could have used RSVP because of the need to highlight the singular words in a "verbatim" sentence recall task in order to activate them in the lexicon.

At first, since the current experiment does not revolve around participants having to recall a sentence verbatim, like in Wellwood et al. (2018), it was thought that verbatim sentence recall was not necessary for the purposes of this study. However, this did not account for the possibility of participants being overwhelmed by the sentence, and thus, cannot perform

the tasks fully. With Rapid Serial Visual Presentation, participants would most likely fully read through each word within the sentence. On the other hand, presenting the sentence entirely within a set time creates a chance that they might not read the sentence entirely before the set time was up. Therefore, it is possible that not using RSVP is detrimental to the experiment procedure because participants were not able to highlight singular words.

4.2.   Limitations and unexpected errors during the experiment

There were some other deviances from Wellwood et al.'s (2018) experiment as well, such as the usage of both native and non-native speakers combined, the location of conducting the experiment (online vs. directly), and how the recall task was conducted (oral vs. re-typing). The deviances were not deemed to affect the "difficulty" of the experiment, but nevertheless, they potentially have consequences towards the results not being as predicted. Of these, the recruitment of participants with both a native English background and bilinguals with English could be a major reason for the deviancies from the predictions. In order to reach as many participants as possible, this current study deviated by recruiting bilingual speakers of English in addition to English monolingual speakers. The 5-year-usage required for bilingual speakers of English was used as a way to theoretically ensure that the bilingual speakers at least know and experienced the language for not a brief moment. Most of the research regarding Comparative Illusions that I have seen so far only employs monolingual speakers. Furthermore, Leivada (2021) showed that there are differences in the performances of monolinguals and bilinguals in terms of detection of illusions as well as in the speed of acceptability judgments.

Additionally, a major part of the stimuli design relied on the two-by-two factorial design. As previously shown in Tables 1 and 2, these two lists were designed to counterbalance each other in terms of the independent variables within the stimuli. While the experiment program did contain two lists with counterbalanced properties, an error in coding made it so that only List 1 (cf. Appendix A) was distributed to all 22 participants. While test trials were conducted by myself, the main preoccupation at the time was to ensure that the program was running and the data could be collected, thus what the program distributed was not looked at. With the absence of the counterbalanced design, it is also not possible to determine whether the effects (or lack thereof) in the Results section were the result of the independent variables themselves or merely because of the order that which the stimuli conditions were introduced.

4.3.   Future research potential

While the hypothesis was not borne out, the actual experiment was not the most desirable either. I will first detail the more practical matters that were not possible for the current experiment that would otherwise be very beneficial. After that, I would discuss the possible applications to improve the participant experience without compromising elements of the previous experiment.

The current experiment was conducted purely online. Willing recruits were sent an email from the experimenter containing the subject number and the link to the experiment. This does have the advantage of allowing participants to do it on their own time, and therefore, it would be more flexible for them. However, this also meant that there was no control over the participants in terms of making sure the deadline for the data collection period was met. Of the 30 people that agreed to participate in this study, only 23 responses were recorded by the end of the collection period. While these 23 responses were partially only achievable in the first place because of this way of data collection, a more preferable method, should another experiment be conducted with a similar methodology, would be to test directly with the participant.

In addition, a lot of the participants' backgrounds were very varied, especially in their age range from 18 to 40. This experiment did not control different age groups because previous literature so far made little mention of the effects of age upon the acceptability judgments. This could be an avenue for future testing regarding the perception of Comparative Illusions. However, research on the processing of sentences describes the older participant group as young as 50 to as old as 88 (Waters & Caplan, Angwin et al.). Therefore, because this age range is still considered only adults, this factor is not expected to matter much to the participants. Furthermore, as previously stated, Leivada (2021) showed that there are differences in the performances of monolinguals and bilinguals in terms of detection of illusions as well as in the speed of acceptability judgments. Therefore, it would be ideal if the bilingual element was not introduced at all in this experiment.

With most of the expression of difficulty from participants surrounding the stimuli sentences being too long, an obvious method of improvement for future research is to reduce the amounts of words within the stimuli. However, most Comparative Illusions are normally around that 11-to-17-word margin already, therefore it is unlikely that that approach is feasible. An alternatively more valid approach is increasing the amount of time allotted to the stimuli. As previously mentioned, the actual time that was allotted for stimuli presentation was actually shorter in this experiment than in Wellwood et al. (2018) (3000ms vs. 2750ms - 4250ms). Additionally, introducing practice trials with grammatical comparatives in future research is advisable as well. This was at first considered to be unnecessary because it was thought that it already would be too long for the participants. In hindsight, it could be a very good way of easing the participants into the tasks that they must do and the length of the sentences without actually recording responses.

5. Conclusion

This study tried to find whether or not actively producing Repeatable Verb Phrases within a Comparative Illusion would affect participants' judgments of the illusion. Following this, the experiment followed a two-by-two factorial design in the stimuli, with the two

independent variables including either the inclusion or exclusion of a Repeatable Verb Phrase and the use of either a probe question task or a sentence retyping task. Each trial was conducted by first exposing participants to stimuli sentences for a short length of time. This was followed by tasks that participants must do either an affirmative or negative probe question or a recall task. After either of the tasks, they have to perform an acceptability judgment task. The dependent variable was the error rates and was determined in target stimuli by the amount of "yes" judgments within the acceptability judgment task. The results were not normally distributed. It showed no significant effect of the inclusion or exclusion of self-production tasks on the acceptability of Comparative Illusions. However, agreeing with the results of Wellwood et al. (2018), there was a significant effect of repeatable Verb Phrases on the acceptability of CIs. Most importantly, there was no significant effect of their interactions. In other words, there was no evidence that people fall for Comparative Illusions more when they were actively producing those Comparative Illusions with verb phrases in which the events within it could be repeated.

6. Bibliography

Angwin, A. J., Chenery, H. J., Copland, D. A., Cardell, E. A., Murdoch, B. E., & Ingram, J. C. (2006). Searching for the trace: The influence of age, lexical activation and working memory on sentence processing. *Journal of Psycholinguistic Research*, *35*(1), 101-117.

Christensen, K. R. (2016). The dead ends of language: The (mis) interpretation of a grammatical illusion. *Let us have articles betwixt us–Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood. Aarhus: Aarhus University*, 129-160.

De Swart, H., & Zwarts, J. (2009). Less form–more meaning: Why bare singular nouns are special. *Lingua*, *119*(2), 280-295.

King, A. P., & Eckersley, R. (2019). *Statistics for biomedical engineers and scientists: How to visualize and analyze data*. Academic Press.

Leivada, E., Mitrofanova, N., & Westergaard, M. (2021). Bilinguals are better than monolinguals in detecting manipulative discourse. *PloS one*, *16*(9), e0256173.

O'Connor, E. (2015). Comparative Illusions at the Syntax-Semantics Interface (Doctoral dissertation, University of Southern California).

Orth, W., Yoshida, M., & Sloggett, S. (2021). Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(6), 906.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633-654.

Waters, G., & Caplan, D. (2002). Verbal working memory capacity and on-line sentence processing efficiency in the elderly. In *Constraints on language: Aging, grammar, and memory* (pp. 107-135). Springer, Boston, MA.

Wellwood, A., Pancheva, R., Hacquard, V., & Phillips, C. (2018). The anatomy of a comparative illusion. *Journal of Semantics*, *35*(3), 543-583.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011, May). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 143-146).

7. APPENDIX A: Target and Filler Stimuli

Table 4: Full list of target stimuli used within the study

List 1 Target Sentences	Self-product ion task present	Repeatab le	List 2 Target Sentences	Self-production task present	Repeatabl e	Probe sentence
As many stargazers saw the 2009 lunar eclipse as the astronomer did.	-	-	As many stargazers saw shooting stars in 2009 as the astronomer did.	-	+	There were stargazers who saw shooting stars.
As many aunts attended Jane's funeral as my uncles did.	-	-	As many aunts attended Jane's dinners as my uncle did.	-	+	There were aunts who attended Jane's dinners.
As many talent scouts attended the 2011 championship game as the coach.	-	-	As many talent scouts attended the 2011 home games as the coach did.	-	+	There were talent scouts who attended the home games.
As many accountants at the firm went bald in their fifties as my lawyer did.	-	-	As many accountants at the firm went golfing on their weekends as my lawyer did.	-	+	There are accountants at the firm who went golfing.
More pop stars got their bellybuttons pierced before the photo shoot than the supermodel did.	-	-	More pop stars got their nails painted before their photo shoots than the supermodel did.	-	+	There were pop stars who got their nails painted before the photo shoots.
More young people finished reading War & Peace this month than old men did.	-	-	More young people resumed reading War & Peace this month than old men did.	-	+	There were young people who resumed reading War & Peace.
More relatives went to my 15th birthday party than my friends did.	-	-	More relatives went to my childhood birthday parties than my friends did.	-	+	There were relatives who went to the birthday party.
In 2002 as many governors ran for president as our senator did.	-	-	In 2002 as many governors praised the policy as our senator did.	-	+	There were governors who praised the policy.
Last spring more ducks hatched from eggs than my chicken did.	+	-	Last spring more ducks laid eggs here than my chicken did.	+	+	There were ducks who laid eggs here.
More businesses in our town burned down in the fire than our school did.	+	-	More businesses in our town held drills after the fire than our school did.	+	+	There were business in the town that burned down in the fire.

More guitar players went deaf from performing live shows than the pop star did.	+	-	More guitar players sold merchandise by performing live shows than the pop star did.	+	+	There were guitar players that sold merchandise by performing live shows.
This year more managers were fired from the company for poor performance than our assistants were.	+	-	This year more managers were rewarded by the company for good performance than our assistant was.	+	+	There were managers who were rewarded by the company.
This New Year's, as many motorcyclists died in accidents as the drunk driver did.	+	-	This holiday season, as many motorcyclists got in accidents as the drunk driver did.	+	+	There were motorcyclists that got in accidents
Last night more passengers in coach were bumped from the flight to Chicago than the businessman was.	+	-	Last year more passengers in coach were bumped from the flights to Chicago than the businessman was.	+	+	There were passengers in coach that were bumped.
More lawyers retired to Florida last year than the judge did.	+	-	More lawyers vacationed in Florida last year than the judge did.	+	+	There were lawyers who vacationed to Florida.
Last season, more American tennis players won their early matches than the Canadian player did.	-	+	Last season, more American tennis players won their final match than the Canadian player did.	-	-	There were American tennis players who won their early matches.
More financial analysts hired from Harvard's business school than our manager did.	-	+	More financial analysts graduated from Harvard's business school than our manager did.	-	-	There were financial analysts hired from Harvard's business school.
More students at my school used Facebook than my teacher did.	-	+	More students at my school joined Facebook than my teacher did.	-	-	There were students at your school who used Facebook.
More senior citizens had their moles removed than our teenager did.	-	+	More senior citizens had their appendix removed than our teenager did	-	-	There were senior citizens who had their moles removed.
More businesses were affected by the new rail system than my house was.	-	+	More businesses were demolished for the new rail	-	-	There were businesses that were affected by

			system than my house was.			the new rail system.
More strawberry plants were watered during the drought than my geranium was.	-	+	More strawberry plants were ruined during the drought than my geranium was.	-	-	There were strawberry plants that were watered.
According to the insurance company, more people with high blood pressure got medical exams than the diabetic did.	-	+	According to the insurance company, more people with high blood pressure got Alzheimers disease than the diabetic did.	-	-	There were people with high blood pressure who got medical exams.
This semester, more girls ate pizza than boys did.	+	+	This semester, more girls wrote their first haiku than boys did.	+	-	There were girls who ate pizza.
This week more bloggers discussed the details of the celebrity divorce than our newspaper did.	+	+	This Sunday more bloggers broke the news about the celebrity divorce than our newspaper did	+	-	There were people who discussed the celebrity divorce.
Last spring, as many houses were cleaned professionally as my apartment was..	+	+	Last June, as many houses were painted professionally as my apartment was.	+	-	There were houses that was cleaned.
Last year, as many freshmen were late to class as the senior was.	+	+	Last year, as many freshmen were expelled from school as the senior was.	+	-	There were freshmen who were late to class.
More botanists photographed the new species in the Amazon than the entomologist did.	+	+	More botanists discovered the new species in the Amazon than the entomologist did.	+	-	There were botanists who photographed the new species.
More quarterbacks were interviewed after the 2011 NFL draft than our wide receiver were.	+	+	More quarterbacks were selected in the 2011 NFL draft than our wide receivers were.	+	-	There were quarterbacks who were interviewed.
This year, more convenience store clerks bought jackpot lottery tickets than my customer did.	+	+	This Friday, more convenience store clerks claimed jackpot lottery winnings than my customer did.	+	-	There were convenience store clerks who bought the lottery.

	Self-production task present	Repeatable Verb Phrase		Self-production task present	Repeatable Verb Phrase	Probe question
Last year, more engineers traveled to San Francisco than our accountant did.	+	+	Last fall, more engineers relocated to San Francisco than our accountant did.	+	-	There were engineers who travelled.

Table 5: Full list of filler stimuli used within the study

Filler Sentences	Self-production task present	Repeatable Verb Phrase	Filler Sentences	Self-production task present	Repeatable Verb Phrase	Probe question
Last spring more ducks hatched from eggs than chickens did due to the declining population of local predators (DP-comp)	+	-	Last spring more ducks laid eggs here than chickens did due to the declining population of local predators. (DP-comp)	-	+	There were more ducks that laid eggs here.
Last fall, more engineers relocated to San Francisco than elsewhere. (CP-comp)	+	-	Last year, more engineers traveled to San Francisco than elsewhere. (CP-comp)	-	+	There were more engineers that traveled.
As many aunts attended my sister's wedding as uncles did (DP-comp)	+	-	As many aunts attended my sister's dinners as uncles did. (DP-comp)	-	+	There were aunts that attended the sister's dinners.
Last year more passengers in coach were bumped from the flights to Chicago than first-class passengers were. (DP-comp)	-	+	Last night more passengers in coach were bumped from the flight to Chicago than businessmen were (DP-comp).	+	-	There were passengers in coach that were bumped.
This year more managers were rewarded than were reprimanded by the company. (CP-comp)	-	+	This year more managers were fired than were promoted by the company. (CP-comp)	+	-	There were managers that were rewarded.
More taxi drivers criticized the union last year than praised it. (CP-comp)	-	+	More taxi drivers joined the union last week than left it. (CP-comp)	+	-	There were taxi drivers that criticized the union.
*As many on their weekends accountants went	-	+	* As many in their fifties accountants went bald at the	+	-	There were accountants who went golfing.

golfing at the firm as my lawyers did.			firm as my lawyers did			
* More football players were visit the Hall of Fame this year invited to than the baseball player was	-	+	* More football players were the Hall of Fame this year formally inducted into than the baseball player.	+	-	There were football players who visited the Hall of Fame.
* Last season, more American tennis players than won the Canadian player did their final match.	+	-	*Last season, more American tennis players than won the Canadian player did their early matches	-	+	There were American tennis players who won their early matches.
More married couples than the bachelor did bought their first house in the suburbs	+	-	* More married couples than the bachelor did looked for new houses in the suburbs.	-	+	There were married couples who looked for houses.

8. APPENDIX B: Error Rates collected from serialised participants in all four testing conditions

Table 6: Error rates from individual participants, with average error rates and standard deviations of test stimuli and filler stimuli of each category

Participant (serialised)	RepVP	Selfprod	Error rate	Average error rate and standard deviation of test stimuli by category	Average error rate and standard deviation of filler stimuli by category
1	-	-	0.38	0.41 (SD = 0.31)	0.54 (SD = 0.38)
2	-	-	0.13		
3	-	-	0.25		
4	-	-	0.13		
5	-	-	0.00		
6	-	-	0.88		
7	-	-	0.88		
8	-	-	0.13		
9	-	-	0.63		
10	-	-	0.71		
11	-	-	0.63		
12	-	-	1.00		
13	-	-	0.13		
14	-	-	0.25		
15	-	-	0.25		
16	-	-	0.25		

17	-	-	0.63		
18	-	-	0.25		
19	-	-	0.13		
20	-	-	0.88		
21	-	-	0.13		
22	-	-	0.38		
1[1]	+	-	0.29	0.48 (SD = 0.38)	0.29 (SD=0.40)
2	+	-	0.00		
3	+	-	0.43		
4	+	-	0.00		
5	+	-	0.33		
6	+	-	1.00		
7	+	-	1.00		
8	+	-	0.43		
9	+	-	0.71		
10	+	-	1.00		
11	+	-	0.86		
12	+	-	1.00		
13	+	-	0.14		
14	+	-	0.14		
15	+	-	0.00		
16	+	-	0.43		
17	+	-	1.00		
18	+	-	0.00		
19	+	-	0.43		
20	+	-	0.14		
21	+	-	0.29		
22	+	-	0.86		
1	-	+	0.14	0.33 (SD = 0.31)	0.20 (SD = 0.33)
2	-	+	0.00		
3	-	+	0.29		
4	-	+	0.00		
5	-	+	0.00		
6	-	+	0.86		
7	-	+	0.86		
8	-	+	0.14		
9	-	+	0.14		
10	-	+	0.50		
11	-	+	0.57		
12	-	+	0.86		
13	-	+	0.14		
14	-	+	0.14		

[1] All 22 participants when this condition category was tested were presented with only *Ungram stimuli

15	-	+	0.29		
16	-	+	0.14		
17	-	+	0.86		
18	-	+	0.00		
19	-	+	0.14		
20	-	+	0.57		
21	-	+	0.14		
22	-	+	0.57		
1	+	+	0.50	0.46 (SD = 0.24)	0.39 (SD = 0.35)
2	+	+	0.00		
3	+	+	0.75		
4	+	+	0.25		
5	+	+	0.43		
6	+	+	0.88		
7	+	+	0.63		
8	+	+	0.38		
9	+	+	0.63		
10	+	+	1.00		
11	+	+	0.63		
12	+	+	0.75		
13	+	+	0.25		
14	+	+	0.25		
15	+	+	0.25		
16	+	+	0.38		
17	+	+	0.38		
18	+	+	0.25		
19	+	+	0.38		
20	+	+	0.50		
21	+	+	0.38		
22	+	+	0.25		