MORE PEOPLE UNDERSTAND ESCHERS THAN THE LINGUIST DOES: THE CAUSES AND
EFFECTS OF GRAMMATICAL ILLUSIONS

By

Patrick Kelley

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Linguistics—Doctor of Philosophy

2018

ABSTRACT

MORE PEOPLE UNDERSTAND ESCHERS THAN THE LINGUIST DOES: THE CAUSES AND
EFFECTS OF GRAMMATICAL ILLUSIONS.

By

Patrick Kelley

A grammatical illusion can be defined as a sentence that seems acceptable, but structurally, the sentence is ungrammatical. Grammatical illusions provide a challenge for linguists to understand why we do not immediately reject illusions like we do for most ungrammatical sentences. One type of illusion that has stirred several ongoing debates is the Escher Sentence. This dissertation focuses on the source of the illusory effect, or the reason why people fail to consistently reject these sentences. This dissertation explores the properties of Escher Sentences, the reason why they are illusory in nature, and what this contributes to our understanding of the parser. Six Experiments were designed to test the acceptability judgments, interpretations, and neurophysiological responses to these sentences. I conclude that Escher Sentences are recognized by the parser as ungrammatical, but because of the structure of these sentences, the parser is tricked into using a coercive operation to force Escher Sentences to have an acceptable interpretation. Escher Sentences gives us potential insight into the constraints of the parser in processing language while at the same time highlighting the parser's strategies in resolving computations that are ungrammatical.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Alan Beretta, for the years of coaching and encouragement that helped me become a better person and a better scientist. He pushed me to always make a better product and to resist resting on my laurels. I've also learned so much from the time we spent outside of Wells Hall, from Iceland to Paul Revere's, it was truly an adventure.

I would like to thank Karthik Durvasula, who was instrumental in helping me learn the statistical and analytical tools I needed to conduct my research and to help round out my skillset as a linguist. He often went out of his way to help with my coding late at night and oftentimes last minute, and without his help, graduate school would have been far scarier.

I would like to thank Marcin Morzycki and Yen-Hwei Lin for being incredibly supportive and informative throughout all my defenses and academic work. They pushed me to think more deeply about theory and how to better incorporate a theoretical approach into my experiments.

I would also like to thank my advisors from my undergraduate experience, Sam Epstein and Jon Brennan. They brought me into linguistics, and they've inspired me to pursue language study now for almost ten years.

Thank you to all my coworkers and linguists who have spent countless hours with me talking about life, linguistics, and everything in-between. Thank you, Alicia Parrish, Kaylin Smith, Drew Trotter, Joe Jalbert, Greg Johnson, Cara Feldscher, Monica Nesbitt, Alex Mason, Matt Savage, Ye Ma, Yan Cong, Tyler Roberts, and Patty Smith.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0   Introduction

In an age of technology where our lifestyle has become automated, it is easy to forget that humans are incredibly efficient machines. In real time, we can derive complex meaning from simple acoustic or visual cues. Even in a loud, noisy, room, we can distinguish between the sounds of various footsteps, music, and other distractions from what we consider the sounds or sights of language. In doing this, we prove to be far more effective at assigning meaning to the noise around us than any computer or home device to date.

As effective as we are, we still are still ineffective in understanding how we accomplish such an impressive feat. This dissertation will provide insight into the limitations of language processing system through the examination of a phenomenon known as grammatical illusions. Grammatical illusions are sentences that sound well-formed, though they are ungrammatical. These are like visual illusions, which fool the onlooker into seeing a shape or image that is impossible, moving, or that simply does not exist. Like visual illusions, grammatical illusions provide impossibilities to a listener or reader, but somehow, we can construct meaning from nothing. Visual illusions are a tool in the visual sciences to test the limitations of our visual system, and this project aims to use grammatical illusions in the same way.

I will explore several facets of a specific kind of grammatical illusion known as Escher Sentences; in the field, they also have been called 'Comparative Illusions' and 'Russia Sentences'. These sentences are highly important for the field for multiple reasons. First, Escher Sentences do not behave like ungrammatical sentences, in that their acceptability judgments and interpretations are unlike typical ungrammatical sentences.

To clarify, grammaticality refers to a binary system, where something either obeys the rules of the grammar (grammatical) or does not (ungrammatical) and acceptability refers to someone's intuitive judgments about a sentence (Schütze & Sprouse, 2014). When looking at ungrammatical sentences in linguistic inquiry, participants are likely to reject them. For instance, people typically rate sentences like *the man saw the dog* as perfectly acceptable and **dog saw the man the* as completely unacceptable. Grammatical illusions, specifically Escher Sentences, elicit highly variable judgments: participants cannot consistently call these sentences acceptable or unacceptable. This variability in judgment makes Escher Sentences particularly fascinating and challenging to study. By studying this variability, we will see that it shines further light on detailing how the parser, or the mechanism responsible for processing language, is able to process sentences, and why it can be fooled. Moreover, syntactic and semantic manipulations directly affect how likely someone is to accept an Escher Sentence, which not only bridges the large gap between theoretical and experimental linguistics but furthers our understanding of how the parser utilizes the tools of the grammar.

Second, through an examination of neurophysiological responses using electroencephalography (EEG), I suggest the parser recognizes that Escher Sentences are ungrammatical sentences. When describing Escher Sentences, there have been claims that the parser is somehow overlooking or ignoring the grammatical constraint or rule that is broken by Escher Sentences or is chalked up to further inquiry (O'Connor, Pancheva, & Kaiser, 2012; Phillips, Wagers, & Lau, 2011; Wellwood, Pancheva, Hacquard, & Phillips, 2017). I will show in this dissertation event related potential (ERP) evidence that there is processing associated with behavioral costs previously observed at a target area in the

Escher Sentence (O'Connor et al., 2012). Specifically, this ERP is associated with recognition of an unexpected stimulus or change in attention as well as indexing discrimination processes which I will extend to early recognition of the Escher Sentence's source of ungrammaticality (Luck et al., 1994; Mangun & Hillyard, 1991; Vogel & Luck, 2000). If the parser realizes that there is an error with these Escher Sentences, it begins to clarify why these sentences behave the way that they do.

Furthermore, it is important for the field as a whole because it helps to clarify the relationship between the grammar and the parser. There are two main hypotheses for understanding the relationship between the grammar and the parser. The first is the one-system view, where the parser and the grammar are one in the same (Phillips, 2013), and the second is the two-system view, where the parser and grammar are separate entities (Ferreira & Patson, 2007). While each camp has their own merits, the question of what kind of relationship exists between parser and grammar is an important question to answer as we work towards creating models of how language processing occurs. This dissertation will aim to clarify arguments between these two camps, though I do not presume to completely discredit one or the other.

Finally, the experiments in this project will show that the parser has inherent biases towards event-interpretations. In sentences where there is an ambiguity as to whether to interpret a particular group as individuals or as an event, it is the case that a majority of the time we side with the event interpretation. For example, in a sentence like *300 dogs ran through the park,* there is an ambiguity about whether there were 300 individual dogs that ran through the park, or if there were just 300 events of some number of dogs running through the park. This bias is important for Escher Sentences, for part of the reason that

they are so illusory is due to an ambiguity of events versus individuals. Because the parser

is biased to interpret these kinds of ambiguities as events, and because of the structure of

the Escher Sentence as we will explore in Section 1.2, it adds up to the variability in

responses. Not only does this project contribute to the understanding of grammatical

illusions, but to the nature of the parser itself.

1.1   Importance of Illusions

Illusions are key in understanding the limitations of a system. When trying to understand how far a system can be pushed, whether with a computer or a human system, scientists will use strange test cases to see how the system responds. Scientists in the visual sciences have used optical illusions for years to understand the limitation and abilities of the human visual system. Arguably, the same logic can be applied to grammatical illusions, where they can be used to test the limitations and abilities of the human language processing system.

Visual illusions can exist because what we see is not a perfect reflection of reality. Optical or visual illusions are often purposefully designed to fool the brain into viewing a stimulus in one way or another, but sometimes they are discovered accidentally. For example, in Figure 1, scientists have crafted what many call The Devil's Fork, an object that cannot physically exist. In the visual landscape, when we encounter such an object, we can construct a shape that our brain will view as some kind of trident or fork (Bach & Poloschek, 2006). At first glance, this appears to be nothing more than a drawn fork or trident, but upon further examination, one will quickly notice that the middle prong in a three-dimensional landscape is connected to nothing. The question here is then why are we able to view this image as a feasible object at all, or, how are we able to visualize an impossible object? This image highlights two important features of the brain: one, that our minds are incredibly powerful and can construct images from shapes outside the realm of physics, and two, that our minds are easily fooled into doing so. Consider a more famous

example that entered the mainstream a few summers ago[1], shown in Figure 2 (BBC News, 2015).

Figure 1 – The Devil's Fork or Devil's Trident, a famous optical illusion where the middle prong comes from nothing.

Figure 2 – The Blue-Black Dress; pictured here is the original undoctored photo.

This picture was taken the summer of 2015, and what started as an innocuous picture between friends quickly became an internet viral phenomenon. In the original

---

[1] Photo credit to Alana MacInnes and Caitlin McNeill, friends debating the color of a dress.

photo above, there is a picture of a dress in the foreground with a natural lighting source to the left. While the dress in question here is blue with black stripes, what people saw when this picture was spread varied greatly. For many viewers, around 57% from a sampled pool of 1401 subjects in a recent study, this dress appeared to be blue with black stripes, but for around 30%, the dress appeared to be white with gold stripes (Lafer-Sousa, Hermann, & Conway, 2015). And for others, blue with brown stripe. There have even been some observations made where the colors may shift for individuals after testing (Lafer-Sousa et al., 2015). To be clear, this dress is actually blue and black, and whatever color variation seen above, whether accurate to the actual dress or not, is a constructed reality from the mind.

The source of why this dress is so puzzlingly prismatic may be attributed to variations within the visual system. Individuals may perceive certain colors based on some internalized preference, though it is unclear whether these preferences are purely mental constructs or physical differences between individuals (Lafer-Sousa et al., 2015). However, scientists were able to shift the perception of the image by manipulating the illumination cue, in this case, whether the background was cooler or warmer in color, indicating that the "illusion" could be flipped on and off (Lafer-Sousa et al., 2015). While visual scientists have been aware of visual illusions for some time, this dress brought visual illusions into the limelight, as there is something inherently exciting about taking a group of people, staring at the same image, and everyone seeing something different.

Considering this information regarding visual illusions, I would like to suggest that Escher Sentences, the focus of this dissertation project, are the *blue-black dresses* of language. The parallel of excitement is not quite internet legendary, but linguists have been

pondering over these sentences for quite some time. Like the dress, there is great variation

in how Escher Sentences are accepted (or viewed). And, like the dress, this variation can

disappear with some linguistic manipulation, meaning that the ability for this sentence to

fool someone can be shifted.

1.2   The Escher Sentence

The star-player of this dissertation is the Escher sentence. Escher sentences provide a problem akin to the problem visual illusions present: data that seem plausible are instead impossible or ungrammatical objects. I first encountered Escher sentences in a classroom my junior year of college, though they first entered the discourse in the acknowledgement section of a dissertation, noted as "the most amazing */? sentence I've ever heard" with its utterance credited to Hermann Schultze (Montalbetti, 1984, pg. 6). Even upon first hearing this, linguists were unsure whether to mark this with an "*", signaling a syntactic violation, or "?", signaling a questionable judgment. Since then, while psycholinguistics as a field has continued to develop, these sentences remain a mystery. The well-known example, thanks to Hermann Schultze, is shown in (1).

(1)   More people have been to Berlin than I have [been to Berlin][2].

Here, linguists propose that when encountering these sentences, participants find sentences like (1) acceptable though in reality there is no available grammatical interpretation (Phillips et al., 2011; Wellwood et al., 2017). At first glance, it seems as if this sentence is a description about how people have gone to Berlin more times than I, the reader, have. However, filling in the ellipsis or deletion site with *been to Berlin,* the sentence becomes much odder: people have been to Berlin more times (e.g. three), than I

---

[2] Brackets throughout this dissertation will represent ellipsis or deletion, so the original sentence would simply be, *more people have been to Berlin than I have.*

have been to Berlin (True/False). Clearly, any computation system cannot compare true and false statements, or propositions, with some number.

The temptation here is to say that people ignore the error in this sentence, or that the parser thinks the sentence is "good enough," where some generalized cognitive mental shortcut or heuristic takes over to fill in any gaps of errors (Ferreira & Patson, 2007). However, I will suggest that neither a good enough story or ignoring an error would suffice to explain why Escher Sentences are the way they are. From experimental evidence using EEG performed in this dissertation (see Sections 3.3 and 3.4), the data suggest that the parser is able to recognize this error, so it is not ignoring it. Moreover, the theoretical approach used in this project (see Section 2.2.4) suggests that instead of relying on a heuristic or mental shortcut to fix the sentence, the parser can access an operation that is available in the grammar, even though usage of this operation is unlicensed.

Nevertheless, Eschers are a key component in understanding how we process information in real time, as the reported variability in the judgment of these sentences begs the question, why do these ungrammatical sentences behave so differently? When participants encounter ungrammatical sentences, they typically find them unacceptable. But, in this case, for many speakers, Escher Sentences are much more acceptable than a word salad sentence like (2).

(2)   *Dragon withered village slay burned the to ground.

Immediately, this sentence jumps out as confusing and an unacceptable sentence of English and is clearly ungrammatical. However, as discussed, (1) is also ungrammatical, so

what makes it different? And specifically, what rules of the grammar does a sentence like (1) actually break? In answering these questions throughout my project, I will also be tackling the bigger theoretical issues at stake, namely, the relationship between the parser and grammar, the idea of an independence between syntax and semantics, as well as the nature of parsing strategies.

1.3   Goals and Structure of the Thesis

The broad strokes of my thesis paint a picture of how grammatical illusions fit into the linguistic landscape, and how they affect some hotly debated linguistic concepts. In this endeavor, the comparison between the value of visual illusions and grammatical illusions is useful, as both push the boundaries for what is possible in a visual and linguistic processing story. Like visual illusions, there are several types of grammatical illusions, which I will cover in Section 2.1. As such, while they share some common ground, they embody a rich diversity of how each illusion works and how strong of an illusion they are. Hence, this dissertation will use one type of grammatical illusion, the Escher Sentence, as a starting point for the exploration into grammatical illusions. As we expand our work into understanding the nature of grammatical illusions more, there will be room for change and further study. For now, the robust ability of the Escher sentences to so easily fool many will serve as the perfect starting point.

The following chapters will break down as follows. First, Chapter 2 will detail the theoretical foundations of the Escher sentence, including multiple approaches to why they are ungrammatical, how they are structured, how they are interpreted, and where the source of variable judgment originates. We will conclude that Ambiguity and Ellipsis Theory (AET) will provide the solution for how Escher Sentences work. Following the theoretical background, Chapter 3 will outline six experiments that put Escher Sentences to the test. These experiments were designed to establish how participants rate Eschers in a variety of settings, how they are interpreted, and what kinds of behavioral and neurophysiological responses participants have. Moreover, Chapter 3 explores the source of the illusion, fine tuning how the effect of being ungrammatical but acceptable works, and

what drives the brain to recognize an issue and variably respond to it. Finally, Chapter 4

will lean on the foundations and lessons gleaned from experimentation to make theoretical

claims about the limitations and abilities of the parser, the relationship between the parser

and the grammar, as well as the properties of Escher Sentences and their illusory strength.

2.0   Background Information

This section covers all the background information needed to understand the content of this dissertation. It will offer a literature review of the topic at hand, though it should be noted that since this is a relatively new line of inquiry for linguistics, there will only be a handful of authorities on some of the specific topics of grammatical illusions. I will also detail how Escher Sentences fit into the big picture issues in linguistics and cognitive sciences.

First, I will walk through the topic of grammatical illusions, breaking down specific examples, specifically negative polarity items (NPIs), attraction errors, and center embedding. Then, I will talk in detail about the syntactic and semantic structure of Escher Sentences, what illusory effects are shared with other grammatical illusions, possible explications for their ungrammaticality, and how we attempt to resolve them. I will conclude the discussion on the nature of Escher sentences with a background into ellipsis and ambiguity, as these are the crucial aspects that make Escher Sentences illusory. Following this, I will fit Escher Sentences into the big picture concepts in linguistics and cognitive science, including how Escher Sentences weigh into the debate between a one- and two-system view, the relationship between syntax and semantics, as well as parsing strategies used when encountering an Escher Sentence. I will conclude this section with an overview of reading electroencephalography (EEG) and event related potentials (ERPs) that will play an important role in establishing the causes and effects of grammatical illusions.

2.1   Grammatical Illusions

Grammatical illusions are a new entry into the arena of linguistic inquiry. As there has yet to be a large body of research developed on this topic, their existence is proving to be both useful and frustrating for linguists. As noted, Escher sentences are part of a larger class of sentences known as grammatical illusions (Phillips et al., 2011). Grammatical illusions are sentences that violate grammatical constraints or rules in the grammar. To elucidate, an example of a grammatical constraint would be syntactic agreement between a subject and verb. If the constraint is maintained, then a sentence like *I like tacos* is grammatical, as the subject *I* and the verb *like* agree with each other. On the other hand, a sentence like **I likes tacos* would violate this grammatical constraint. In the case of illusions, a constraint is violated, but instead of eliciting a consistent judgment, participants lack consistent intuitions of acceptability. This lack of consistency is known as selective fallibility (Phillips et al., 2011).

While this selective fallibility has been observed in many different grammatical illusions, including attraction errors, negative polarity items, center embedding, etc., it is unclear if there are connections to be made within each type of illusion with respect to how selective, meaning if there are greater processing costs comparatively, or if there is some way to quantify how selective a sentence would be. The notion of selective fallibility appears to be a feature of grammatical illusions rather than an explanatory concept. I will argue in this dissertation that selective fallibility is not imposed on an individual basis, but rather, there is evidence to suggest that this fallibility lies in the very structure of the illusion and can vary depending on the sentence.

Setting aside momentarily the issue of selective fallibility, there are two other challenges posed from grammatical illusions that lie ahead for linguists to solve. Firstly, there is a need to create a model of grammatical illusions that clearly explains why, given our strong language capabilities as humans, some violated grammatical constraints go largely unnoticed. While I will spell out a subset of these illusions in the following sections, each illusion has different properties and therefore different etiologies for their selective fallibility, which leads to the second problem of grammatical illusions, creating a model that unifies illusions. In other words, is there a common feature that each grammatical illusion has that leads to their selective fallibility? Classifying what kinds of errors are susceptible to illusory behavior, meaning that they are not clearly and consistently judged by participants, is a challenge (see Phillips et al., 2011), and while a satisfactory answer to this question has yet to be found, focusing on one particular type of illusion, Escher Sentences, may clarify what issues the parser is encountering.

To clarify, the parser is the mechanism that many linguists attribute to the processing of language. While it is not the case that there is an actual parser somewhere in the brain, as linguists, we model our understating of language processing as if there were some computational piece of machinery that processes sentences in a systematic manner. The parser for human language is quite powerful, as it is able to take in incredibly complex stimuli such as acoustic sound waves (verbal speech) or visual disturbances (sign language) and within milliseconds create meaningful information. If the parser is capable of handling tons of complex operations, then as linguists we are incredibly interested in knowing when it fails. And even so, this notion of failing needs further clarification, as evidence from Sections 3.3 and 3.4 suggest that the parser does not fail to recognize errors,

but it does fail to accurately produce the computational output (a crash) and instead making an interpretation work.

Nevertheless, as part of this endeavor to understand parsing and grammatical illusions, I will walk through three other well-known illusions outside of Escher Sentences: negative polarity items, attraction error sentences, and center embedded sentences. These other grammatical illusions will provide further insight into how the parser handles these odd cases. Though all these grammatical illusions share some features, it is clear that each illusion operates differently in the syntax and, given recent findings, elicit different behaviors and neurophysiological responses as well. For each illusion, we will discuss what constraint is being violated, what participants believe they are reading (or how they fix the problem), and how selective fallibility is affected.

2.1.1   Negative Polarity Items

Negative Polarity Items (NPIs) refer to lexical items that must be licensed with a

special context in order to be grammatical (Uribe-Echevarria, 1994). For example, consider

the usage of *ever* below in (3).

(3)   a. No dog with an electric collar ever ran away.

         b. *A dog with an electric collar ever ran away.

In (3a), the adverb *ever* works with the sentence, as the negative context of *no dog*

licenses its usage. In this case, we would say that the negation in (3a) licenses or allows the

usage of *ever*, which is the lexical item referred to as the NPI. Contrast this with (3b),

where there is no negative context, and *ever* seems weird to use in this context. Therefore,

we could hypothesize that any item that is an NPI would require a negative context.

However, this is not always the case. For the past fifty years there have been massive

efforts to create a model explaining how NPIs work (Drenhaus, Frisch, & Saddy, 2005;

Lasnik, 1972; Sohn, 1995; Uribe-Echevarria, 1994; Vasishth, Brussow, Lewis, & Drenhaus,

2008), yet these efforts have been met with great challenges from theorists and

experimentalists alike. As hypothesized, the data in (3) suggest any NPI requires some

form of a negative context to license or allow its usage, but now consider a sentence like

(4).

(4)   Who has ever seen a dog run away with a collar on?

With the sentence in (4), the usage of *ever* seems to be okay in this context, yet there is no negative context that can technically license it. If we assume that every NPI must be licensed, then then how do we explain the acceptability of a sentence like (4)? While theories that detail the nature of what is licensing a sentence like (4) (see Giannakidou, 2011 for a summary), much work has also been done in the experimental realm. The experimental findings of testing NPI sentences are quite surprising: rather than having clear-cut judgments for sentences with licensed NPIs in (3) and unlicensed in (4), participants were very inconsistent. This would suggest that some NPI sentences are behaving like grammatical illusions. Specifically, a timed grammatical judgment task, replicated with an ERP study, revealed what the authors called an "intrusion effect," where a licensor that was structurally inaccessible had participants rarely judging these sentences as ungrammatical (Drenhaus et al., 2005). To elaborate, an example sentence of this illusory effect can be seen in a sentence like (5).

(5)   ?A dog who chewed on no sofa was ever a bad boy.

In (5), the negation here, albeit an uncommon negation in English, is in an embedded clause and cannot c-command *ever,* yet participants are truly bad at rating this sentence poorly, as replicated again in an eye-tracking study (Vasishth et al., 2008). In this case, the *no* does not enter a c-command relationship with *ever* because *no* is embedded inside a relative clause *who chewed on no sofa*. The authors claim that participants were inconsistent with their judgments in a sentence like (5) because this sentence is a grammatical illusion that violates the constraint of requiring the NPI to be licensed: the

negation in the embedded clause has a feature that cues in the processor to substitute

where a c-commanding licensor should be (Vasishth et al., 2008). In other words, they

claim that the parser inserts a cue where a licensor should be and incorrectly licensees the

sentence. Thus, while NPIs are still hotly debated, NPI sentences seem to fit the bill for a

grammatical illusion, as they violate a constraint determined by the grammar yet are

selectively fallible.

In short, NPIs have the same general behavior that classifies them as grammatical

illusions: one, they violate a constraint in the grammar, and two, this violation goes largely

unnoticed, which arguably means that it is selectively fallible. Moreover, given that this

result was replicated using different paradigms, this increases the likeliness that these

results are robust. The next section will go over another type of illusion, attraction errors,

that elicit these same general characteristics.

2.1.2   Attraction Errors

The second type of grammatical illusion discussed here are sentences with

attraction errors. Attraction errors occur in sentences where a determiner phrase (DP)

"attracts" the subject-verb agreement incorrectly, usually due to the structural closeness of

the attracting DP. Let us first consider an example in (6).

(6)   ?The password for the computers are on the hard drive.

Here, the DP *the computers* is "attracting" the agreement of the verb *to be*

immediately after the DP. The verb *to be (are),* is looking to enter into a subject-verb

agreement relationship, and *the computers* are the closer potential partner both

structurally and linearly in the sentence. So, an attraction error occurs, and *are* is

morphologically realized as opposed to the grammatical *is.* Reading these sentences in

isolation makes this particular error fairly easy to identify compared with NPIs.

Nevertheless, repeated experiments suggest that these types of agreement errors in

experimental settings, where participants are typically tasked with answering questions or

are being asked to pay close attention, consistently go unrecognized by the participant

(Bock & Miller, 1991; Eberhard, Cutting, & Bock, 2005). In other words, participants in

these experiments are failing to see that *the cabinets are on the table* is an ungrammatical

continuation in this sentence. While this error may seem trivial, meaning that participants

still can understand what the sentence means, their acceptability judgments go unaffected.

A linguist might predict here that participants would notice the grammatical error, still

understand the gist of the sentence, and mark it lower in acceptability, usually on a Likert

Scale of 1-7 or 1-5 (Likert, 1932) where the higher the number, the more acceptable the sentence is. This, however, is not the case, as participants consistently rate sentences like (6) quite highly.

It is also important to note here that with respect to attraction errors, while selective fallibility refers to missing a violation and believing a sentence is grammatical, there appears to be a lack of evidence for a participant to miss a violation or constraint and believe a sentence is ungrammatical (Phillips et al., 2011). To clarify, consider our previous example in a grammatical form—

(7)   The key to the cabinet was on the table.

There is a lack of evidence with respect to attraction errors to suggest that participants will undermine the value (7) and call it ungrammatical, which provides an interesting question for selective fallibility: why can we be tricked into thinking something is grammatical but not ungrammatical? One argument may be made, in Section 2.1.4, where we may be tricked into thinking something is ungrammatical with grammatical center embedded sentences. In these cases, this may be due to processing overload or memory issues rather than grammaticality. Nevertheless, it is interesting that grammatical illusions seem to only be operating in one direction i.e., ungrammatical sentences fool us into thinking they are good but not grammatical sentences fooling us to think they are bad.

What makes attraction errors like (6) even more puzzling is that while the human parser has trouble with simple errors such as agreement attraction, we are much more likely to notice complex errors (Phillips et al., 2011). In a sense, missing a word or two in a

sentence may be considered trivial as in a sentence like (6), though sometimes a single word can obfuscate a sentence. One way to think about this would be to consider errors that involve movement, an operation that would involve complex processing on the parser's part (Phillips et al., 2011; Wellwood et al., 2017). Movement means that words in the sentence are shifting structurally, creating gaps where words once were. In linguistics, years of experiments and theory show us that these gaps where words have moved from are important for our grammar. To understand this, let us construct a pair of sentences where these gaps can greatly affect our judgment of a sentence, shown below in (8).

(8)   a. Who did Randy believe that Sheryl liked ___ ?

        b. *Who did Randy believe the idea that Sheryl liked ___ ?

To begin, (8a-b) show a pair of sentences with movement and gaps that look very similar but are wildly different with respect to their grammaticality. In (8a), let us assume that Sheryl likes Ricardo, and Randy believes that Sheryl likes Ricardo. One could form a question about this situation in (8a) by asking who the person was that Randy believes that Sheryl likes. In English, when formulating this kind of question, our syntax performs a movement operation that moves Ricardo from the "___" and moves it to the front, replacing him with the wh-word *who.* In syntactic theory, this "___" is known as a gap and can be represented in a number of ways, depending on the theoretical framework. Now, for a sentence like (8a), this movement of Ricardo to the front of the sentence is typically allowed, but sometimes the addition of one or two words can drastically affect the sentence, as shown in (8b). Here, the same movement of Ricardo is occurring, but instead

of moving out of a simple statement like *Sheryl liked Ricardo,* he instead moves out of what

we call a complex DP, in this case, *the idea that Sheryl liked Randy.* This type of movement

or extraction, as some linguists call it, out of a complex DP is not allowed, and this is clear,

as the acceptability of (8b) is markedly lower than its counterpart in (8a).

This type of violation is called an island violation. Islands are syntactic structures

that do not allow for extraction or movement out of them, and in this example, this is a

classic island violation known as a complex noun-phrase (NP)[3] island (Ross, 1967).

Specifically in this case, the wh-word, originally Ricardo, moves out of a definite DP,

eliciting an ungrammatical judgment (Ross, 1967; Szabolcsi, 2006), and from a

psycholinguistic perspective, this type of sentence yields an increased processing load for

participants (Hofmeister & Sag, 2010).

From all this movement and talk of islands, it is amazing that we as humans can see

these types of complex errors and almost immediately recognize them as troublesome

sentences, but some simple agreement errors go, to the best of our knowledge, unnoticed.

Thus, attraction errors exemplify another type of grammatical illusion, where a

grammatical constraint, subject-verb agreement, goes unnoticed and is selectively fallible.

The next section will cover the last grammatical illusion background known as center

embedding.

---

[3] While there are data to suggest that there are languages with DP and languages with NP (Bošković, 2005), regarding English, most linguists agree that DPs represent what many used to call NP for many years (Abney, 1987). Therefore, I will be using DP instead of NP for English sentences throughout this dissertation.

2.1.3   Center Embedding

This section covers the third type of what may be considered a grammatical illusion given our definition. Center embedding was first noted in 1956 as not having any particular limit to how far grammatically we could take it, though grammatical and processable are very different beasts (Chomsky, 1956). Seven years later, Chomsky and Miller provide this sentence which is often used in linguistics classrooms around the country (1963):

(9)   The rat the cat the dog chased killed ate the malt.

This construction, even for a trained linguist, is hard to understand even after breaking it down clause by clause. For clarity, this sentence is referring to a dog that chased a cat, a cat that killed the rat, and the rat that ate the malt. While some have tried to put a number to the limit of how many center embeddings there can be e.g. three written, preferably none spoken (Karlsson, 2007), some have also noted that when you remove one of the verbs at the end of a sentence like (9), which is technically ungrammatical, participants report these as just as acceptable as the grammatical version with three verbs (Gibson & Thomas, 1999). Arguably, center embedded sentences can be considered a grammatical illusion because they fail to elicit a clear judgment and there is some variability in responses. However, the question here is exactly what grammatical constraint is violated. While this may be a constraint unexplored in English grammar, this could also be due to memory limitations or some other cognitive function that is failing here. What is clear, though, is that the ungrammatical version of (9), missing an external argument, is considered as acceptable as (9). Hence, center embedded sentences fits the bill for being

36

called a grammatical illusion: an ungrammatical sentence where acceptability judgments indicate selective fallibility.

        While we have explored just three of the more well-known grammatical illusions, the focus of this project is on the Escher sentence, which I will discuss in the next section. Importantly, I will also reframe the argument around selective fallibility, in that it is not necessarily the case that the brain is missing an error, but rather that the responses to the error itself invoke variability not on an individual level, but with respect to the parser's abilities.

2.2   Escher Sentences

        The fourth and final type of grammatical illusion and star-player outlined in this

chapter is the infamous Escher Sentence. As mentioned in the introduction, the Escher

Sentence was first noticed, to the best of the field's knowledge (Phillips et al., 2011;

Wellwood et al., 2017), from a dissertation by Montalbetti (1984), where he mentions

Hermann Schultze as giving him the infamous sentence in (10).

From Montalbetti, 1984

(10)   More people have been to Berlin than I have.

        Regarding this sentence, he then states, "[s]ome have taken this sentence to be a

proof of the autonomy of syntax!" (pg. 6). This is part of an ongoing debate that deals with

how sentences are processed: does the syntax inform the semantics, can the semantics

operate separately, and what kind of relationship does this have with respect to the

information in the grammar vis-a-vis the parser? There have been some that suggest that

the semantics can override the syntax (Ferreira, Christianson, & Hollingworth, 2001;

Ferreira & Patson, 2007), which would assume that the syntax and semantics can operate

autonomously, while others are staunchly against this claim, stating that the parser and the

grammar are one and the same, so there should be no need for autonomous units to

operate outside of the processing system (Phillips, 2013). Now, this issue has a lot of facets

to it, and for this project, I will discuss how Escher sentences can help to clarify at least

some of these questions. I will return to this debate in Section 2.3.1.

From what we know so far, Escher Sentences are another type of grammatical illusion that violates a grammatical constraint that goes largely unnoticed. However, unlike the other illusions, this type of illusion provides some further challenges, as it is not immediately obvious what constraint is being violated. Moreover, its selective fallibility can be easily manipulated to be more illusory in nature, meaning more likely to be overlooked, or less illusory to the point of easily being assigned awful ratings from participants. First, I will walk through ways in which to change the illusory strength of Escher sentences, and following this, I will lay out several possibilities to explain the violation of Eschers and possible fixes that participants are giving it.

2.2.1   Eschers – Shifting Acceptability

Like other grammatical illusions, it is possible to change the acceptability of the illusion by manipulating features of the structure. For a working example, let us use (11) shown below as our exemplar Escher Sentence.

(11)   More dragons flew around the village than the fairies did.

The first manipulations of Escher Sentences to affect acceptability, to my knowledge, came from a poster presented in 2004 where researchers found that the ellipsis at the end of the sentence, in this case *the fairies did (fly around the village)*, was important to the acceptability of the sentence (Fults & Phillips, 2004). For them, having elided content made the sentences more acceptable than if there was no ellipsis, i.e., no words were deleted at the end of the sentence. While there were a few issues with this study, namely comparing blank space with content, I will address this effect later in Chapter 3. Nevertheless, this led to others finding ways to shift the acceptability of the Escher sentence, with the next big project in 2009 that found that if the predicate of the comparative clause was repeatable, and that if the subject of the comparative clause was plural, the sentences were markedly more acceptable (Wellwood, Pancheva, Hacquard, Fults, & Phillips, 2009). To see this, consider (12).

(12)   a. More dragons flew around the village than the fairies did.

b. More dragons burned the village than the fairies did.

c. More dragons flew around the village than the fairy did.

Between (12a) and (12b), the predicate has been shifted from a repeatable one in (12a), flying around a village, to a non-repeatable one in (12b), burning a village down. As reported in Wellwood et al.'s results, there was a distinct drop in acceptability from sentences like (12a) to (12b) (Wellwood et al., 2009). Similarly, from (12a) to (12c), the subject in the comparative clause goes from plural to singular which also yielded a drop in acceptability (Wellwood et al., 2009). For my research, I will also show that these separations create a standard gradation between Escher Sentences. Escher Sentences that have plural DPs are referred to as Strong Eschers, and Eschers with singular DPs are Weak Eschers. Strong refers to how "strong" the illusory effect is, meaning that people will more likely fall victim to the illusion, and "weak" refers to an Escher that is more likely to be rated lower or have a weaker illusory effect.

These manipulations reveal more deeply what goes wrong with an Escher sentence and what participants are thinking when they interpret and accept them. I will now walk through three theories that detail both what the error is and what the possible solution may be.

2.2.2 "How Many" Operator Theory

Escher Sentences appear to satisfy the definition of a grammatical illusion, as they violate a grammatical constraint as well as induce selectively fallibility. While the selective fallibility of these sentences is well-known across several studies, it is unclear what grammatical constraint is violated. One approach is to suggest a violation of syntactic dependency. Consider a pair of sentences in (13), comparing an Escher (13a) with a grammatical comparative in (13b).

(13) a. ?More Brazilians ate hamburgers than the Americans did [eat hamburgers].

       b. More Brazilians ate hamburgers than Americans did [eat hamburgers].

The difference between (13a) and (13b) is the presence of the overt determiner in the comparative DP, rendering (13a) odd. Intuitively, one may look at the properties of *the* as the perpetuator of the Escher effect (Abney, 1987; Heim, 1982; Heim & Kratzer, 1998), though having an overt determiner does not necessarily induce an Escher as shown in (14).

(14) The Brazilians ate more hamburgers than the Americans did.

While the structure of the overall sentence has changed, if *the* were truly driving the effect, then (14) should be an ungrammatical or at least an Escher Sentence, but it is neither, instead being a plain grammatical comparative sentence. Instead, the theory here from Wellwood et al. focuses on a "how many" operator theorized to be covert in English that requires a bare plural (Bresnan, 1973; Wellwood et al., 2017). Bare plural means a

plural DP without an overt determiner. A model of this approach is shown in an example

pair of sentences in (15).

(15)   a. *More Brazilians ate hamburgers than *how many* the Americans did.

        b. *More Brazilians ate hamburgers than the *how many* Americans did.

        c. More Brazilians ate hamburgers than *how many* Americans did.

    The *how many* in this case makes a clear difference between (15a-c), as this

operator requires a bare plural, so feeding it a determiner like in (15a-b) results in an

ungrammatical sentence. To further this point, the English covert operator is overt in

languages as the Bulgarian *kolkoto.* Consider this Bulgarian sentence set from (Wellwood

et al., 2017):

(16)   a. Poveče amerikanci sa bili v Rusija ot-kolkoto slonove sa bili v Rusija.

        More   Americans   are been to Russia from-how.many elephant.PL are been in

        Russia

        "More Americans have been to Russia than elephants have been to Russia"

        b. ... *ot-kolkoto az /slon-ăt /slonove-te.

        from-how.many I /elephant-the/elephant.PL-the

        "than I/the elephant/the elephants"

    To start, (16a) is the sentence in Bulgarian, spaced out to best match the English

direct translation below it. From here, a sentence in quotation marks is given as the most

accurate translation given English syntax. So, when reading this data set, it is important to pay attention to the * mark in the Bulgarian line in (16b). Here, the comparatives *than I/the elephant/the elephants* are ungrammatical, following the same pattern in English. To elaborate, when *kolkoto* is given the equivalent of anything that is not a bare plural noun, the sentence is ungrammatical.

Therefore, if it is the case that a *how many* operator exists in English, this operator would create a wh-dependency between the matrix and comparative clause over degrees of individuals; this dependency is then violated by the inclusion of any DP that is not a bare plural (Wellwood et al., 2017). Wellwood and company then go on to run a series of experiments with the idea of the ungrammaticality of the Escher Sentence coming from not satisfying this operator. They then conclude that an event-interpretation is the repair strategy used to resolve this issue, where the matrix clause, because it has an event-interpretation available, licenses an event comparison even though the syntax does not support it (Wellwood et al., 2017).

While this approach works with the data, and though it posits additional covert structure, it is unclear if the data collected in Bulgarian reflects the same types of selective fallibility as do Escher Sentences in English. This is because there is a lack of experimental evidence with these judgments. While the judgments given may very well be correct, because these are grammatical illusions, it is imperative to see the spread of variability in judgments. Without data supporting the variability in the judgments of these sentences, then it is arguable that (16b) is not a grammatical illusion but just an ungrammatical sentence. Furthermore, though they claim that the event-interpretation strategy is preferred, there is little motivation as to why it is appealing. While I agree that participants,

for the most part, view Escher Sentences as event comparisons, I will refine their claims by

proposing an explanation as to why this interpretation is easily allowed by the parser. I will

also assume that this syntactic approach does not fully encapsulate the data available for

Escher Sentences. The next section details another hypothesis to explain why these

sentences are ungrammatical.

2.2.3   Comparable Sorts Theory

While the "how many" operator theory places the onus of the error on the syntax,

another alternative is that the grammatical constraint being violated is one of comparable

sorts in the semantics. In every Escher Sentence, *more* is always the first word, and this

sets us up for a clear semantic comparison to be made—

(17) More (A, B) =1 if |A| > |B|

(17) states that more A than B is true (=1) if the cardinality of A is greater than the

cardinality of B. In most cases, this formula works for comparative sentences in English.

For example, consider (18).

(18)   More Brazilians ate sandwiches than Americans did.
        |10| |5|

Here, whether we assume the sentence is referring to an event of eating a sandwich,

or the number of people involved, the formula in (17) here works. In (18), there are 10

Brazilians or 10 events of them eating sandwiches which is greater than either 5 Americans

or 5 Americans eating sandwiches. It is also the case that the comparative portion could be

interpreted as a proposition, but because the comparative DP is a bare plural, the

comparison between events or individuals is preferred. However, with Escher Sentences, if

we were to try to apply the formula in (17) to one, there would be a problem. (19)

demonstrates the problem encountered by the Escher Sentence when trying to compute a comparison.

(19)   More Brazilians ate sandwiches than the Americans did.

|10| T/F

In (19), if we complete the rest of the sentence, it states that there are more Brazilians or Brazilians eating sandwiches (expressed as some amount, in this case, 10 times or individuals) than the Americans ate sandwiches, which is a true/false statement or proposition. Therefore, this crashes the semantic interpretation formula provided in (17). Structurally speaking, the Escher Sentence does not appear to have any problem syntactically, but the semantics has a problem computing the comparison trying to be made. If it is the case that in the matrix clause, or the first clause before the comparative portion, is interpreted as either events or individuals given the scope[4] of *more*, then there are serious problems comparing this with a comparative clause dealing in truth values. Logically, no system of grammar in any language can compare numbers with truth values.

Whether we adopt a hypothesis like the *how many* approach*,* or if we adopt a more semantic approach, we can safely conclude that this sentence is indeed ungrammatical. But, the puzzle is nowhere complete, as both the *how many* and *comparable sorts* approaches leave out crucial aspects of the Escher Sentence. Hence, I will provide two additional hypotheses that explain the qualities of the Escher Sentence.

---

[4] Taking scope is a term used by semanticists that describes when a lexical item or operator can rule or help define relationships with other words in the sentence.

2.2.4 "Who" Insertion Theory

Another theory that could possibly explain why participants are arriving at an interpretation would be an assumption about syntactic insertion near the ellipsis site. Simply put, participants may be inserting a "who" silently into the structure to yield a grammatical sentence. If participants are inserting a *who* immediately after the comparative DP, then the sentence may be interpreted as grammatical, because even filling in the ellipsis site fully is masked by the *who* creating another relative clause. A syntactic tree[5] is shown in Figure 3 that would detail where this insertion may be occurring.
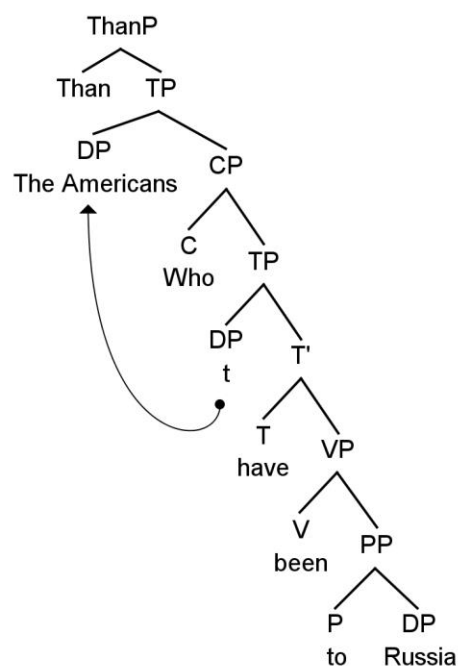
Figure 3 – Partial syntactic tree of where "who" may be inserted.

---

[5] Than Phrase (ThanP) is used here for *than* (see Wellwood, 2015).

In this structure, a relative clause is introduced after the comparative DP, which forces the ellipsis to resolve to a grammatical conclusion i.e. more Brazilian folk have been to Russia than the Americans who have been to Russia. Now, there is a grammatical comparison between two groups, the number of Brazilian people going to Russia versus the number of Americans who have also gone to Russia. Furthermore, no matter the theoretical approach to ellipsis, whether there is either partial/full syntactic/semantic information, the relative clause makes it only possible for this sentence to be grammatical. While this approach offers a simple take on Escher Sentence processing, it assumes that the syntax makes a rather large change to the structure, which would be reflected by a large processing cost. O'Connor and company found processing costs at this target area (2012), and Wellwood and company have found other costs associated with semantic changes (2009), which makes it less likely that the syntax takes over at this point. Moreover, in an EEG experiment, we would expect a P600, or a positivity 600 ms post stimulus or target area (Osterhout & Holcomb, 1992), however from the experiments in this dissertation, this ERP is absent. This solution is an alternative approach to understanding the Escher Sentence, but the theory presented in the next section provides the best story to date for explaining the Escher Sentence.

2.2.5   Ambiguity and Ellipsis Theory (AET)

One of the core components for creating an Escher sentence is to have the matrix predicate contain an ambiguity with respect to an event or individual interpretation. This ambiguity is well-documented in work done by Krifka (1990). Consider a set of his examples shown below in (20).

(20)   a. Four thousand ships passed through the lock last year.

b. The library lent out 23,000 books in 1987.

In (20a-b), there are two possible interpretations for the predicate. The first is an individual or object meaning e.g. in (20a) there were four thousand individual ships passing through the lock, or in (20b), there were 23,000 unique books lent out from the library in 1987. Alternatively, the second possible interpretation for these sentences is an event interpretation. For example, in (20a), there could have been some number of ships that passed through the lock 4000 times, or in (20b) there were some number of books lent out 23,000 times, meaning that some books could be lent out multiple times. This ambiguity is quite prevalent in English, and in later experiments I perform, Experiments 5 and 6 (see Sections 3.5 and 3.6), I will show that we are biased in the case of (20) towards interpreting these as events. Krifka calls this event interpretation event-counting, where individuals are taking this ambiguity and resolving them to events through a mechanism he refers to as *Object-Induced Event Measure Relation (OEMR)* (Krifka, 1990)*.*

Essentially, OEMR is an operation that interprets a DP with a measure value to it, so, in the case of (20), 4,000 ships, and produces a predicate interpretation for events. For

Krifka, he makes an assumption about a null-determiner that does the work for the OEMR (Krifka, 1990). However, this null determiner is unattested, so if we move away from this assumption, OEMR is a coercion operation that takes an entity, a number, and a predicate, and creates an event interpretation. I propose a refinement to the OEMR, which I will use as a theoretical basis for constructing a theory of Escher Sentence interpretations.

I will call this theoretical approach Number to Event Operation (NEO). The idea here is that NEO is a function that takes three arguments, a property of events, a property of individuals, and a number, and from this produces an event interpretation. This operation would be readily available in the grammar as it accounts for sentences like (20) to either have an individual or event interpretation. To exemplify this, consider (21).

(21)   NEO(e)(t)(n)

        e=property of events, t=property of individuals, n=number

(21) provides a more mathematical way to interpret this function. Essentially, if the semantics can calculate *e, t,* and *n,* then the output would be an event interpretation. To use an example from earlier, let us apply NEO to (20) shown below in (22).

(22)   4,000 ships passed through the lock last year.

        NEO(e)(t)(n)

        NEO(passed through the lock)(ships)(4000)

In the case of (22), NEO takes in a value *e passed through the lock,* a value *t, ships*, and a value *n, 4000*, and outputs an event interpretation such that there were 4,000 events of ships passing through the lock last year. Importantly, NEO is an optional operation for deriving a semantic interpretation as this sentence without a NEO operation can be computed as a sentence about 4,000 distinct individual ships.

Ontologically speaking, the ambiguity between events and individuals in sentences like (20) is unique in that other classifications of general entities behave slightly differently. A classic approach to breaking down the difference between an event and individuals, or stages and instantiations, comes from Carlson's dissertation (Carlson, 1977). Following this approach, entities can be broken down thusly—

(23)
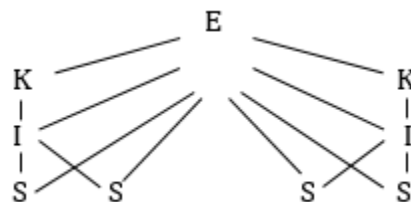
Here, entities (E) represent any entity or individual in the universe. These entities can be divided up into three distinctions, or using the technical term, *sorts.* The first and broadest kind of sort is called a *kind* (K, e.g. *dogs are winners*) which refer to a general type of thing. From kinds, there comes another sort called *instantiations* (I, e.g. *My dog is a winner*), which refers to a specific instance of an individual. Finally, the last sort and most specified one are called *stages* (S, e.g. *My dog hiding under the sofa from the storm last*

night was still a winner) which are specific individuals with temporal and spatial information. Note that this graph is done in a symmetrical manner to reinforce the idea that each sub-level beneath entity are also themselves entities. The motivation for constructing an ontology like this comes from evidence where in certain sentences, only particular levels of entities are allowed. Specifically, some nouns and predicates may only apply to kinds, instantiations, or stages. Consider sentences like (24).

(24)   a. The tyrannosaurus rex is extinct/#The lobster is extinct.

        b. Randall is feeling paranoid/#Dogs are feeling paranoid.

        c. David ate pizza at Papa John's yesterday/#Dogs ate pizza at Papa John's yesterday.

The paradigm in (24) has a set of three sentences, with predicates e.g. *the tyrannosaurus rex is extinct,* that require kinds, instantiations, and stages, respectfully. In the "/" section of each sentence, the DP of the predicate shifts, and the sentence becomes semantically odd. For example, (24c) makes sense for David, representing a stage, with a particular location and time works well, but when trying to fit that predicate with a kind *dogs,* the sentence is markedly worse.

        Returning to (20), then, the linguistic prediction would be that each predicate can only accept one sort of entity. But, as we define what kind of entity *ships* are in (20a), they can either be a stage or instantiation, which goes against our understanding of predicate and entity relationships. However, thanks to an operation like OEMR from Krifka, or a slightly revised version like NEO, these entities can be realized as either stages or instantiations (events or individuals). The ontological distinction between stages and

instantiations is not enough to explain an event-object ambiguity. Most likely, it is a combination of both the properties of the entity and the predicate that lead to an ambiguity. As the parser passes by the predicate *four thousand ships passed through the lock*, it entertains both interpretations. Somewhere further down the parse, if the ambiguity is resolved, it selects events or individuals accordingly, but if the ambiguity is not, an operation to make an event like OEMR or NEO is used instead.

Regarding Escher Sentences, each Escher Sentence has an event-individual ambiguity in its matrix clause, and in the comparative clause, there is no syntactic to semantic information to help resolve whether the sentence is about events or individuals. Wellwood et al. conclude that what they call the *event comparison hypothesis* is what drives the "Comparative Illusion Effect" (2017). Here, they say that the illusion of the Escher Sentence can exist because events are a readily available interpretation and is quite tempting for the participant even though it is unlicensed (Wellwood et al., 2017). It is unclear why this event-reading is so tempting given Wellwood et al.'s story.

While I agree with their claim that participants are interpreting Escher Sentences as event comparisons, I propose a deeper explanation as to why they are so "tempted" to do so. First, let us assume that the parser itself is biased to select for events in this ambiguity cases, which means that we are more likely to use an operation like NEO in a sentence like (20). While evidence in Section 3.5 and 3.6 point directly to this bias, a bias by itself still does not fully explain why Escher Sentences resolve into an event comparison. Given the specification of NEO, in an Escher Sentence, there is no n value, or number, to plug into the formula. One reason the parser may misapply NEO and fill in a number would be due to an ellipsis that potentially could contain number information. In other words, because there is

an illicit ellipsis site where what is being elided may be unclear, it leaves the parser in a

position where it does not have number information available. Furthermore, if it is the case

that event interpretations from ambiguous contexts are frequent, this suggests that an

operation like NEO may be used frequently. If so, the parser uses NEO for the Escher case,

coerces a number, produces an event comparison, and thus computes an acceptable

interpretation.

To elaborate, the second component in creating an illusory effect is to have an

ellipsis site in the comparative clause, specifically verb-phrase or VP-ellipsis. VP-ellipsis

outside of the context of Eschers looks like (25).

(25)   William wanted to play the piano, and Yujin [wanted to play the piano] too.

Here, the elided or deleted material is bracketed, and for many speakers, this type of

ellipsis is quite natural, as it seems odd to reiterate information recently discussed in the

discourse (Rooth, 1993). Amongst syntacticians, however, there is still considerable debate

as to what information is being elided. There are several ways to conceptualize what

Merchant calls, "meaning without form" (Merchant, 2013, pg. 1). This debate can be

summarized into two questions. First, what kind of content does the ellipsis site contain?

Once words are deleted, what information is left behind is still unclear. Second, what is the

nature of the relationship between the ellipsis site and antecedent? There are a few

possible answers here. Some theorize that the ellipsis site contains syntactic information

(Merchant, 2001, 2013; Sag, 1976) , where others have argued there is no syntactic

structure in the ellipsis site (Culicover & Jackendoff, 2005). Additionally, there is a divide

between the identity between antecedent and ellipsis, where some believe it to be strictly syntactic (Sag, 1976), some semantic (Culicover & Jackendoff, 2005), and some others claim that there is both syntactic and semantic information (Kehler, 2002; Merchant, 2013).

Regarding the theoretical divides in the field, I will assume an analysis where there is complete identity between the antecedent and the ellipsis site, with the ellipsis site containing both syntactic and semantic structure. However, though the outcome of any theory of the ellipsis is the same for Eschers, as in it does not work, what information is inside the ellipsis site could potentially impact what claims this dissertation can make. Nevertheless, consider (26).

(26)   More knights have seen the queen in her robes than the squires have [seen the
        queen in her robes].

(26) is another example of an Escher Sentence with elided material in brackets. If we assume full identity, this means we assume that the deleted material is the exact same structurally as it was in the antecedent. But, even if the ellipsis site contained something else, e.g. some truncated version, the problem remains. Theoretically, if the matrix clause has an ambiguity in event vs. individual interpretation, the ellipsis site would as well. However, this is not the case, as the Escher comparative clause is a proposition. To push this a little further, even if we assume that only the word *seen* is kept in the elided material, this is still a proposition and not an event: more knights have seen the queen in her robes (six knights/six times) than the squire has seen (True? False?). On top of this problem, no

matter what the parser may fill into the elided space, it still has to deal with the fact that

the matrix clause is ambiguous and the comparative clause does little to help.

To try and understand then how the parser can take an Escher Sentence and

somehow create meaning from it, we need to look more closely at the ellipsis site. The first

piece of experimental evidence we have deals with filling in the ellipsis site completely.

When this is done, the acceptability of the sentence degrades significantly (Fults & Phillips,

2004). As a quick experiment, try to read (27) out loud without any ellipsis, and it is likely

to be much worse in acceptability than (26).

(27)   ?More knights have seen the queen in her robes than the squires have seen the

        queen in her robes.

For Fults and Philips, they arrived at the conclusion that (27) is worse off than a

sentence like (26) using acceptability judgments. However, their comparisons were made

between elided material and blank space. Therefore, any effect found by Fults and Phillips

could be that the longer sentences were marked as less acceptable than shorter ones, or

that redundancy of repeating material affects the judgment (Bresnan, 1973; Rooth, 1993;

Wellwood et al., 2017). While it is likely the case that sentence length is only a partial

feature for the decline in acceptability, this will be addressed in this dissertation in

Experiments 5 and 6.

Nevertheless, it appears that by removing the ellipsis, or having a sentence like (27),

the sentences behave more predictably i.e., they are rated lower for being ungrammatical.

This suggests that ellipsis is an important piece of the puzzle in understanding why Escher

Sentences behave the way they do. I propose that the very presence of this ellipsis contributes to the mistake of using NEO on a sentence that does not fulfill its requirements. Because this ellipsis does not create a grammatical sentence as ellipsis typically does, it is considered to be unlicensed. However, because this ellipsis does exist for the Escher Sentence, it makes room for the parser to "fill in" the required information it needs to arrive at a successful parse. Specifically, the parser is tempted to fill in a number, *n*, for the comparative clause because an unlicensed ellipsis *could* have that information, even though structurally it does not. The presence of this illegal ellipsis, combined with a bias towards making events, tricks the parser into inserting a number for the comparative clause, using NEO to make it into an event comparison, and makes a sentence like (26) to mean something like *more knights have seen the queen in her robes (three times today) than the squire saw her in her robes (one time today).*

Regarding the parser, as it goes through and processes a sentence like (26), it is unclear at what point the ambiguity of event versus individual is resolved. However, previous research points to processing costs after the auxiliary verb (O'Connor, Pancheva, & Kaiser, 2012; Wellwood et al., 2017). This processing cost may amount to a decision by the parser to choose an interpretation for the matrix event-individual ambiguity, but as I will suggest in Chapter 3, these costs may be associated with the recognition of the illegal ellipsis site. Therefore, ellipsis is important for the illusory process: if the parser fills the Escher Sentence ellipsis site with extra information to allow for an event interpretation, then the illusion has succeeded. If instead the parser crashes, the illusion effect has failed. Importantly, without ellipsis in play, there can be no illusory effect.

2.2.5.1 AET specifications

I have created a theoretical approach that attempts to address each problem of the

Escher sentence. I call this theory *Ambiguity and Ellipsis Theory (AET),* and this will be the

framework used as the basis for our understanding of Eschers. AET can be summarized in

(28).

(28)   Ambiguity and Ellipsis Theory (AET)

Escher Sentences are ungrammatical because the comparative computation

between matrix and comparative clauses is incomputable. This is further

complicated with a comparative clause with an unlicensed ellipsis that does not

resolve the matrix ambiguity between events and objects, leaving the parser in an

ambiguous state. However, because the parser is inherently biased to interpret

these ambiguities as events, and because the ellipsis site provides enough further

ambiguity to entertain the possibility of including number information, the parser

utilizes the Number to Event Operation (NEO), that resolves the sentence. Because

the Escher Sentence can vary in structure and thus vary in how illusory it is, NEO

may or may not be used. This variability is what yields in many cases an event

interpretation and in some a computational crash.

First, we covered two different possibilities to explain the ungrammaticality of

Escher sentences, a syntactic approach from Wellwood and company that posits a covert

how-may operator (2017), and a semantic one where the comparison being made is

between events/individuals and a proposition. Because the semantic theory of *comparable*

sorts does not postulate the presence of an unknown operator, and because it fits into a larger general framework of explaining every iteration of an Escher Sentence, this most likely is the source of its ungrammaticality. Furthermore, on top of the ungrammaticality of the Escher Sentence, the reason it is illusory, or variable in judgments, is due to a combination of an unresolved ambiguity of events and individuals in the matrix clause, a bias towards events, and an unlicensed ellipsis. Let us work with a specific example to detail how AET and the NEO operation would play out with an Escher Sentence, demonstrated below in (29).

(29)   More knights have seen the queen than the squires **have** [seen the queen].

At the point of the bolded verb *have*, the parser is now faced with an interesting problem: what goes inside the elided structure? Clearly, if it fills it in with full identity, we have a huge problem i.e. an ungrammatical sentence. So, being biased towards events, and by the nature of the parser wanting to ascertain meaning as much as possible, one possible way it could find a usable interpretation from an Escher Sentence like (29) would be in (30).

(30)   More knights have seen the queen than the squires have.
        Apply: NEO(e)(t)(n)
        $\exists$n NEO(seen the queen)(squires)(n) $\land$ n>max{n'|n'-many knights seen the queen}

In (30), NEO is looking for three inputs, but cannot find an n value in the comparative clause. So, a possible way that the parser responds to this would be to use an n value here where n is greater than the maximum value of n' such that n' is the number of times knights have seen the queen. In other words, as long as the knights have seen the queen more times than the squires, whatever number that may be, then NEO assigning this value here yields a computable comparison between the number of events of knights seeing the queen and number of events of squires seeing the queen. This strategy is appealing to the parser because 1) the parser likes events and 2) the ellipsis of the Escher Sentence does not do the job ellipsis typically does i.e., remove material to make a sentence grammatical, so the parser assumes something has gone wrong and fills in information. However, because the parser is acting based off no grammatical support, this operation will sometimes fail, and it will fail more often when the comparative DP is singular or the predicate is repeatable as evidenced by acceptability data (O'Connor et al., 2012; Wellwood et al., 2009, 2017).

2.3   Big Picture Issues – Impacts on the Field

This section will discuss the field of linguistics and psycholinguistics more broadly and how the research into Escher Sentences may impact it. First, I will talk about what parsing means and in general how we think about a mental grammar, which will lead into the common divides of a one- vs two-system view. This will lead into a discussion about the nature of syntax and semantics, and how they are related to each other with respect to language processing accounts. Then, I will briefly discuss the concepts of serial and parallel processing, concluding with a discussion on the nature of parsing strategies.

2.3.1   Parser, Grammar, and Syntax/Semantics

To return to the statement made by Montalbetti, Escher Sentences may provide

some evidence that the syntax and the semantics operate separately (Montalbetti, 1984).

To the psycholinguist, this statement raises a few concerns: first, with respect to the parser,

does the parser have access to all the grammatical information it needs, a syntax,

semantics, phonology, etc., which is called a one-system view, or instead does the parser

rely on extra-grammatical features, a two-system view, where the grammar and parser are

different entities. Furthermore, if it is the case that the grammar and parser are distinct

entities, does this also provide grounds for assuming that the semantics and syntax can

operate autonomously, meaning that if the syntax fails, can the semantics step in and vice

versa? And, when we interpret language data, do we proceed serially, meaning one thing at

a time and in a particular order, or parallel, where all kinds of processes are operating in

tandem.

In the one-system view (Phillips, 2013), the parser has complete access to all the

grammatical constraints and rules, which poses a potential problem for illusions: why

would the parser fail to recognize these errors when it has complete access to grammatical

information about them? This is where the strongest argument for the two-system view

steps in, since there is room for this error to live, as the parser, being a separate entity, can

call upon both the grammar and heuristics or mental shortcuts, a framework called "good

enough parsing" (Ferreira & Patson, 2007). In this story, the parser misses the error

because it operates separately from the grammar, and in the case of something like a

grammatical illusion, the parser says "good enough" and works out an unspecified semantic

heuristic to make the sentence work. This is an example, then, of the syntax and the semantics working autonomously from each other.

So, how can Escher Sentences help to clarify this debate? For the one-system proponent, illusions provide the core challenge of being variably judged in the first place, because if the grammar and the parser are one and the same, there should be no problems in detecting any violation. For the two-system proponent, the core challenge is to provide an actual framework for the heuristics being used by the parser as well as providing when these heuristics come into play.

Let us assume that it is not the case that the parser is missing the error in an Escher Sentence. Rather, the parser recognizes something has gone awry, but it is unclear if the ellipsis it has flagged can be rescued. Since the parser is biased towards interpreting an ambiguity of individuals and events as events, it works it out what it is most familiar with and goes with an event interpretation for the whole sentence. In this case, the grammar and parser may be one and the same, since the grammar informs it that there is a problem, and using the grammar, it also assumes the sentence is about events thanks to an inherent bias within the parsing mechanism. This yields an illusion, not due to extra heuristics or separate processes, but by the very nature of the parser and grammar itself. However, this could also still be framed in a two-system view, where the assumption and corrections the parser makes are outside of the grammar.

The concept of the parser being flawed is human, since humans are also flawed beings, so it is not farfetched to imagine that the parser, like the visual system, can be tricked by its own processes. There are still a few fine-grained issues to iron out, such as by what processes the parser uses its event-bias to transform the sentence into something

64

grammatical. Perhaps it uses a relative clause *who,* or perhaps it reorganizes or coerces information into the ellipsis site.

Arguably, without more experimental data to work with, both the one-system and two-system story may still work for illusory data. Regardless of whether they rely on heuristics or on rules of the grammar, these illusory sentences purposefully trick the system. Just as visual illusions can fool an incredibly powerful visual system in our brains, the same goes for these kinds of linguist-made constructions. To my knowledge, it is very rare for a native speaker to naturally produce an Escher on their own, though this neither adds or detracts from the argument. The point here is that these sentences are purposefully made to subvert, so even though the parser has all the information it needs, it can still be fooled given the limitations of trying to complete an impossible sentence. Nevertheless, once we have concluded the experimental section, I will weigh in again on this debate in Section 4.2.2.

2.3.2 Serial vs. Parallel Parser

One divide in the cognitive sciences that receives a lot of attention to this day is the idea of processing and whether it proceeds in a serial or parallel fashion. Serial processing is when processes are done one at a time and only moves onto the next process when the previous one is complete. By contrast, parallel processing is where all processes are run in tandem (Townsend, 1990). Imagine a computer that is trying to run three commands: open a word processor, turn on music, and open an internet browser. Now, the user may select these three options one at a time, but it is unnecessary to wait for each command to load completely before moving onto the next one, and we could imagine the user writing a script to open all three simultaneously. Regardless, the computer proceeds in a parallel fashion, opening each command as it receives it and without waiting for a particular one to finish, it can initiate opening other commands. This is parallel processing.

Now, on the other hand, imagine a machine line where there are three steps in making a toy: gluing the parts together, inserting the batteries, and painting. It would make sense for the workers to start by gluing the toy together, then inserting any electronics, and finally painting the near completed product: trying to do all three at once could prove disastrous. This is serial processing.

In both cases, scientists have spent the last hundred or so years trying to imagine cognitive processes in similar ways e.g. when we see a stoplight, do we first process the visual stimulus of seeing the light, and only then initiate the chain of cognitive commands that tell us what to do, or rather, are we taking in stimuli and issuing cognitive commands all at once? With respect to the problem set of grammatical illusions, claiming that these errors/illusion are done in a serial or parallel fashion does not impact the theory and work

at hand. However, it is important to note that this debate thrives in every linguistic subfield and strategy.

There appears to be a general consensus in the field that linguistic processing is done in both serial and parallel fashions (Pinker & Prince, 1988). Scientists have sought to tease apart serial versus parallel processing by examining reaction times and accuracy (Townsend, 1990) as well as EEG (Gow, Keller, Eskandar, Meng, & Cash, 2009). The most likely scenario is that the brain uses both parallel and serial processing to accomplish its cognitive processes. Nevertheless, there may be room to further examine what kinds of processes are used when rescuing the Escher Sentence, which may be a useful direction for further study.

2.3.3   Parsing Strategies

When the parser encounters linguistic input, it immediately goes to work processing and assigning meaning to the input. In a perfect world, every sentence uttered would be clear, grammatical, and unambiguous, but this is unlikely. The parser must make myriad decisions as it processes the linguistic input it receives, including gaps and dependencies, transformations, and ambiguities. On top of grammatical constraints on language, physical constraints on mechanically transforming visual or acoustic disturbances into meaning, and working within both memory and time constraints all present challenges to the parser. The decisions that the parser makes in assigning meaning to the input received are known as parsing strategies (Frazier, 1979). One lens to examine these strategies through is by comparing grammatical sentences; for instance, consider (32) (Fodor, 1978)—

(32)   a. The horse raced past the barn fell.

b. The horse raced past the bard and around the duckpond.

Here, (32a), known as a garden path sentence, elicits a high degree of difficulty in processing, where (32b) is markedly easier to digest. By looking at the two of these sentences together, with respect to the ambiguity the parser encounters with *the horse raced...,* there is a preference for interpreting *the horse raced* with *raced* being an active verb of *horse* i.e. the horse does the racing, rather than the passive relative clause i.e. the horse was raced by someone else (Fodor, 1978).

Considering Escher Sentences, I propose a similar bias, in that we are biased to view Escher Sentence matrix predicates as events over objects or individuals because there is a

68

general bias for such an interpretation. However, this bias cannot be the only feature of an Escher Sentence, as the prediction would be that everyone would just view them as event-related sentences, which is not the case. Because ellipsis is present in the comparative clause, a choice the parser can make is to crash or stick with an individual interpretation. With this in mind, it creates a picture of our parser that has inherent biases, which again we can see right here in (32): we are biased towards the simpler construction of an active over a more complex passive relative clause. Hence, Escher Sentences can contribute to an understanding of a parser's inherent biases.

2.4   Introduction to Electroencephalography (EEG) and Event Related Potentials (ERPs)

For the final section of our background tour, I will spend some time discussing the history of EEG and ERPs, what exactly this machinery is measuring and how, and going into some predictions for possible ERPs that we may encounter in this direction of study.

2.4.1   A Brief History

The first recorded usage of EEG was by Hans Berger in 1929, where he used a single electrode to demonstrate that electrical activity from the brain can be recorded (S. Luck, 2005). Over the course of the next decade, EEG came to be more accepted by the scientific community, and following World War II, the usage of EEG for both clinical and scientific research started to grow amidst the cognitive revolution (S. Luck, 2005).

Following this, the first instance of a scientific publication linking a cognitive ability with ERPs was in 1964, when scientists discovered an ERP component they called the *contingent negative variation* (CNV). This describes a negative deflection that signaled a participant's anticipation of a stimulus to come (Luck, 2005; Walter, Cooper, Aldridge, McCallum, & Winter, 1964). CNV sparked a whole new world of working with EEG and cataloging ERPs, and it did not take long for the first language-related ERP to come about, the P300, or P3, in 1965 (Sutton, Braren, Zubin, & John, 1965). In this case, the P300, or a positive deflection 300 milliseconds post stimulus, was elicited when participants encountered a stimulus whose modality, whether auditory or visual, they had failed to predict (Luck, 2005; Sutton et al., 1965). Similar to the CNV, this positive deflection hinted that prediction and anticipation would play an important role in processing, both cognitively and linguistically.

Over the course of the next several decades, several famous ERPs would be well documented and discovered. Two of the most well-known and used in the field today are the N400, a negative deflection 400 milliseconds post stimulus that the authors associated with semantic integration difficulties (Kutas & Hillyard, 1980), and the P600, a positive deflection 600 milliseconds post stimulus that the authors associated with syntactic errors

(Osterhout & Holcomb, 1992). However, this division is not so simple: because technology

increases exponentially, our understanding of even these basic ERPs is muddied as we

begin to confront them with more and more data. We are learning that N400 and P600

does not always equate to semantic and syntactic errors, respectfully, and that these ERPs

can relate to several other linguistic processes, including non-violations (Kutas, Van Patten,

& Kluender, 2006; Lau, Phillips, & Poeppel, 2008), not to mention the fact that P600s can

also occur in musical violation contexts (Fitzroy & Sanders, 2013).

   With this in mind, the study of ERPs continues to be challenging. On the one hand, as

we continue to study the brain and ERPs, we hope to refine our understanding of what

linguistic events these ERPs are indexing, but we are forced to realize that this goal is very

far from fruition. For example, the early characterizations of the P600 as an index of

syntactic violations or of repair of temporary syntactic ambiguities, has been complicated

by the observation that certain types of semantic anomalies can also give rise to a P600

(Kim & Osterhout, 2005). A great deal of current ERP research seeks a more adequate

understanding of ERP components and what elicits them.

2.4.2   Limitations of Electroencephalography (EEG)

The tool of choice for looking at Escher Sentences is EEG. But before indicating that this tool is ideal for the job at hand, it is important to recognize its limitations. EEG records electrical activity of the brain at the scalp, though the scalp is a poor conductor. Not only that, but inconsistency in brain tissue and cerebrospinal fluid mean that the electrical signal is hard to pinpoint an exact location from where it came. That is, ERPs recorded at one electrode could reflect activity that is quite dispersed in the brain. Thus, EEG is not useful as a means of determining where activity is taking place; this is referred to as having poor spatial resolution. If good spatial resolution is important to some research question, then other technologies would be used, most notably functional magnetic resonance imaging (fMRI), positron emission tomography (PET), or computed tomography (CT). These technologies have excellent spatial resolution, or knowing where activity occurs, since we can indirectly track cerebral blood flow to regions where there is never cell activity. This sets up a demand for usable energy which is transported by the circulatory system. However, the technologies that are informative regarding brain location have poor temporal resolution, or knowing when activity occurs, as they are necessarily limited by the speed of blood flow, which is on the order of seconds rather than the millisecond scale that neuronal activity operates on.

EEG is therefore not the best tool to find out where activity occurs in the brain. The advantage of EEG is that it does have millisecond resolution, so it is able to record brain activity literally as the milliseconds go by. EEG is thusinformative about processes that are occurring inside our brains as they happen. This is critical for an understanding of the incredibly fast operations of the parser i.e. neuronal activity occurs in milliseconds.

73

Crucially, our theories of sentence processing also operate on a millisecond scale, meaning that language effects in EEG are very quick. For my work, since I am interested in knowing when linguistic processes occur with respect to grammatical illusions, EEG provides an excellent measure for determining in real time what the parser is doing at any given moment. This methodology will be applied to Escher Sentences in Sections 3.3 and 3.4.

3.0   Experiments

Chapter 3 contains six experiments that explore Escher Sentences more deeply. There are several questions to be answered regarding the nature of grammatical illusions and Escher Sentences. (1) what interpretations, if any, are people constructing when encountering these sentences; (2) what do acceptability judgments look like for Escher Sentences; (3) how does the brain react to Escher Sentences and what types of responses are there; (4) does the timing of experiments affect the behavioral results of Escher Sentences; and (5) what properties can be manipulated to alter the acceptability and interpretations of Escher Sentences?

The experimentation process began in 2012 and concluded at the end of 2017. Experiment 1 was the first study conducted and sought to answer what interpretations people arrive at when reading Escher sentences. This was done by conducting a broad survey using Mechanical Turk and Qualtrics software (*Mechanical Turk*, 2017; *Qualtrics*, 2017). Experiment 2 uses the same survey approach but was done with paper surveys using college students from MSU. This experiment explored the acceptability judgments of Escher Sentences.

Experiments 3 and 4 branch out to using an EEG and acceptability judgment tasks. The judgment tasks was done in tandem when testing with EEG to build upon the acceptability findings in Experiment 2 as well as exploring if there are any neurophysiological responses to Eschers Sentences assuming processing costs found at the site of elision (O'Connor et al., 2012). Out of these EEG experiments, Experiment 3 compares Weak and Strong Eschers directly, while Experiment 4 introduces a baseline comparative control. Experiment 5 further tests the acceptability judgments and

interpretations of Escher Sentences while also testing whether timed or untimed settings affect the results. Finally, Experiment 6 replicates and refines the results for Experiment 5 with the inclusion of an additional control condition. All in all, these data are designed to tackle the larger issues at hand for Escher Sentences, namely, what explains their variability, does the parser recognize that Escher Sentences are ungrammatical, what evidence is there for Ambiguity and Ellipsis Theory (AET) and the nature of the parser, as well as uncovering evidence that add to the debate of a one- versus a two-system parser and grammar relationship.

3.1   Experiment 1 – Mechanical Turk Behavioral Study

Abstract

Experiment 1 was a behavioral study on the interpretations of Escher sentences. 96 participants completed a survey conducted using Mechanical Turk as a crowd-sourcing engine, where they rated comparative statements as true/false/not sure given a context. These statements contained grammatical controls and Escher Sentences. Participants' overall responses suggest they consider grammatical controls and Escher Sentences to be about events. However, individual bivariate correlations reveal participants were tracking individual interpretations with Escher Sentences. This highlights the highly variable nature of Escher Sentence interpretations.

Escher Sentences present a complex anomaly for study, as by the theoretical framework established, there should technically be no interpretation. Yet, upon asking several peers and naïve speakers, especially non-linguists, the common response to a sentence like *more people have been to Russia than I have* is typically a furrowed look, and thoughtful silence. This is typically followed by some jumbled explanation from the speaker that more people have gone to Russia than the speaker has. To then explain to the speaker that this sentence means nothing evokes a wide range of responses, from, "oh, I see" to "no, I still think it's okay."

When I was introduced to Escher sentences, my first question posed to my peers was, "what do these sentences even mean?" From the discussion in Chapter 2, we concluded that there is no available interpretation given the structure of the Escher Sentence. Following AET, there is an ambiguity of events vs. individuals in the matrix clause that goes unresolved with an ellipsis that would yield an ungrammatical sentence if it were

77

filled in. Moreover, at its core, the comparison is flawed computationally, as it tries to match a number of individuals or events with a proposition. While a computer might crash at this moment, humans do not, and instead create and compute some meaning, and what interpretation people arrive at is precisely what Experiment 1 sets out to explore.

Experiment 1 addresses the question of what interpretation people walk away with when they encounter Escher Sentences. Because I wanted to ascertain a general idea of this sentence outside of a circle of classmates and linguists, I opted to use Mechanical Turk (*Mechanical Turk*, 2017) that was touted as a valid approach to linguistic surveying (Gibson, Piantadosi, & Fedorenko, 2011; Sprouse, 2011). Mechanical Turk is a software developed by Amazon that crowd-sources opinions for a project. For example, if you were an advertising firm looking to see if a new would be popular amongst a general population, you could use Mechanical Turk to ask hundreds of workers, or what Amazon calls Turkers, to view a series of images or slogans and rate them. Turkers are paid per survey or questionnaire that they complete, ranging anywhere from $0.05 and above per completion. This allows for scientists or whomever is using the software to draw upon vast subject pools. For instance, in Gibson et al's acceptability study, for example they recruited 519 Turkers, vastly outnumbering the typical subject pool a linguist or psychologist might draw from (Gibson et al., 2011).

Because of the successes from other authors, I believed this method to be best to gather a large number of subjects and to see how they were interpreting these sentences. Due to the variable nature of Escher sentences, it was best to build a foundation from a large population to work from. Moreover, what differs in this experimental context than

plainly asking people what these sentences mean is that they are provided some options to

choose from rather than construct a meaning on their own.

3.1.1   Methods

3.1.1.1 Overview

The primary objective of Experiment 1 was to determine if there were any trends in interpretation judgments across participants. To test for interpretations, the first challenge was allowing participants to express what they were interpreting without leaving a blank space to write something in. This would ultimately lead to a large variety of answers, and as experienced previously, participants find these sentences difficult to interpret and explain what they are. While there may be some interesting directions for memory capacity and short-term memory recollection here, Experiment 1 opts to remove this possible cognitive barrier and provide readers with scenarios to read and choices to make. I designed a series of scenarios, roughly two to three sentences in length, that detailed some event. After each scenario, the participants read a single statement that summarized some claim from the scenario. Upon reading this scenario and paired statement, participants were asked to say if that statement was true or false given the context. They also had the option to respond "not sure" in order to avoid participants forcing an answer if they truly did not have one.

The main idea here was to see how participants interpreted statements that made comparative claims about a particular context. These statements were a mixture of grammatical comparatives and Escher Sentences. I sought to compare how individuals were interpreting grammatical comparatives, and how they were interpreting Escher Sentences to look for any patterns. Importantly, these statements were designed to see if participants were viewing Escher Sentences like comparisons of individuals, comparisons of events, or something else entirely.

This survey[6] was designed using Qualtrics software (*Qualtrics*, 2017), which allowed for more freedom in design and data collection than Turk offered at the time. Qualtrics also had software implementations to ensure that participants answered each question without skipping. For example, if a worker wanted to just click through as quickly as possible, they at least had to respond to each question rather than continuously pressing "next." If participants were found to be doing this i.e. completion times way below the average, they were excluded to ensure that only responses that were fully read and answered were collected. I also collected demographic information, allowing participants to self-report their age, gender, and native-speaker status. These options were left for the user to fill in themselves in order to include as many participants as possible.

Once participants filled out their demographic information as well as entering their Turk ID[7] to ensure compensation, they began answering questions one scenario at a time. Upon completion of the survey, they were given a message saying that they have completed the survey and were immediately compensated through Mechanical Turk's system. All data were kept anonymous thanks to the Turk ID system and having Qualtrics information locked away so only investigators had access to the response data.

3.1.1.2 Stimuli
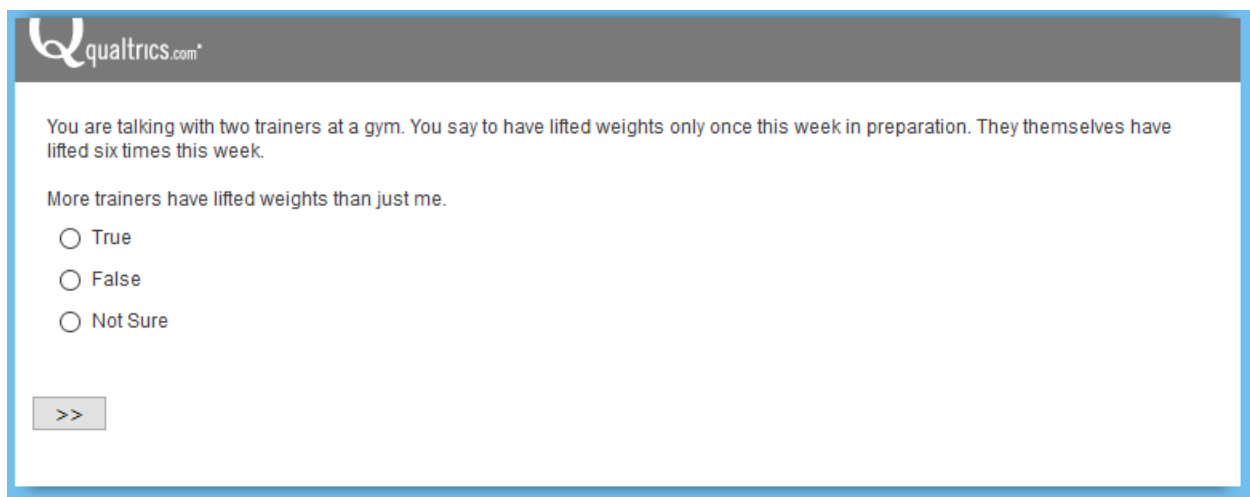
The survey itself was comprised of forty scenarios, each of which had two versions, A and B, which kept the overall structure of the story but changed one vital aspect. In the A scenario, the reader performed some task more than others, and in the B scenarios, the

---

[6]This is an active link to the survey; if you like to sample it, you can enter any string of text into "Worker ID" and continue from there. https://umich.qualtrics.com/jfe/form/SV_87jm03cHQhN4brf
[7] Turk IDs were an anonymous string of characters that allowed for Turkers to complete tasks and ensure compensation through Mechanical Turk.

reader performed some task less. Each of these scenarios was paired with a statement that

reflected a comparison about the events occurring in the scenario. There were three

possible statements. The first kind of statement was called Statement One (S1). This was a

grammatical comparative that directly and unambiguously compared individuals using

"just me" as the comparative DP element. Figure 4 below is a screenshot of an example

scenario paired with a S1 statement.

Figure 4 - Example of survey screen for participant with an S1 statement following an A scenario.

In the scenario from Figure 4, there are two trainers and the reader; after some

discussion, the reader finds out that they lifted weights once while the two trainers have

lifted weights six times this week. The statement below the context, S1, *more trainers have*

lifted weights than just me, would be true, since there are two trainers versus one person,

just me. This was one possible combination that readers could encounter, but they could

have also been given a B scenario, where the reader lifted six times, and the trainers only

lifted once. In that case, with this same S1, the answer here would still be false, since "just me" will always be less than the group of others in these scenarios.

The second possible statement, Statement Two (S2), was another grammatical control, but this time took the form of unambiguously comparing events, as shown below in (33).

(33)   The trainers have lifted weights more than I have.

In the Scenario from Figure 4, if this statement were to follow, the solution would clearly be true, since the trainers have lifted six times and the reader has only done so once. In the other Scenario version, B, where the reader lifts six times instead and trainers only once, this statement would then be false. Finally, the third possible statement, Statement Three (S3), was always an Escher Sentences and took a familiar form—

(34)   More trainers have lifted weights than I have.

Now, given (34) and the scenario in Figure 4, where trainers have lifted six times and the reader only once, it is up to the reader in how they interpret this sentence: is it true, because there are more trainers, or is it true because the reader has lifted only once? The only way to tease these apart is by manipulating the scenario and statement combinations and tracking how they rate S3 with S1 and S2 statements. S3 thus serves as the experimental condition, and S1 and S2 are the grammatical controls. Table 2 below summarizes the predictions for each scenario and statement combination.

Scenario	Statement	Interpretation	Prediction
A	1	Individual	TRUE
B	1	Individual	TRUE
A	2	Event	FALSE
B	2	Event	TRUE
A	3	Escher	??
B	3	Escher	??

Table 1 – Predictions for each scenario/statement combination.

For scenarios that focus on individuals i.e. A1 and B1, the comparisons between others vs. reader were always true since those statements contained "just me" (=1 individual) versus a group of others (>2). In the event cases, A2 and B2, the correct response varied between true and false. Additionally, correct responses contained both true and false answers. This was done to avoid learning effects that may have occurred if the correct response were always true or always false. Furthermore, because a "not sure" option was included, this prevented participants from forcing an interpretation when they may not have had one.

Importantly, if participants answered true for both A3 and B3, this would pattern with the S1 statements, meaning they were relating Escher Sentences to statements about individuals. If participants answered false for A3 and true for A2, this means they were related Escher Sentences to statements about events. If no pattern emerged from answering S3, then this would suggest participants may be guessing and most likely are not

arriving at any interpretation. This would be especially true if participants respond with

not sure for a majority of the S3 or Escher Sentence statements.

      The design of these stimuli also helped to avoid other possible confounds. First,

returning to the scenario in Figure 4, there are always two trainers and yourself, the

reader. However, the reader is never directly told that they themselves are part of that

group i.e., a trainer, which provides the freedom for the participant to either be in the

group or not. For example, in a statement about comparing individuals, the reader can

either be a trainer or not. But, in a statement about comparing events, it is crucial that the

reader is not a trainer for a statement comparing events to be acceptable. The freedom to

not lock the reader into any group within the context prevents the reader from associating

themselves with the groups in the context potentially altering their judgments.

      Another problem could be a potential distributive vs. collective reading. For

example, consider a sentence like (35).

(35)   The trainers have lifted weights once.

      Given that our scenario in Figure 4 has two trainers, a reader could potentially read

this in two ways: one, a distributive reading, where there are two events of each trainer

lifting weights once; or two, a collective reading, where there is a single event of lifting

done by both trainers. Since we cannot control how participants are reading sentences, to

avoid this issue, the stimuli were designed such that if the reader performs some action a

number of times, that number would always be greater than the number of the people in

the scenario. Therefore, regardless of collective or distributive readings, the TRUE/FALSE predictions will still hold.

Finally, when designing these stimuli, it was important to control for bare-subject readings. Escher Sentences using comparative plural DPs require the usage of *the*, so there was not a problem here. However, for the other grammatical comparatives, having a bare plural subject could cause ambiguity. To illustrate, consider a sentence pair like (36).

(36)   a. Trainers have lifted weights.

b. The trainers have lifted weights.

Here, the bare-subject sentence, (36a), without an overt determiner, implies that the set of trainers in general have lifted weights, while the overt determiner DP, (36b), denotes a particular set of trainers. This offers a potential confound for readers, as having something like *trainers* could denote some other group outside of the context. While unlikely for participants to imagine trainers in general given this experiment, in order to avoid any possible confound, every S2 has an overt determiner to clearly indicate that the group the statement is referencing is from the context provided and not elsewhere.

Overall, with forty scenarios, divided into A and B, and three possible statements, there were 120 total combinations possible for a participant to see. These combinations were divided into six equal groups (Surveys 1-6) using a Latin Square Design, each with 20 of 120 possible combinations of experimental stimuli and 20 filler scenarios and statements, totaling 40 questions per participant. The ratio of fillers to experimental items was therefore 1:1. To clarify, a Latin Square Design is commonly used in behavioral

experiments to avoid participants seeing the same stimuli within an experiment. If

participants see a scenario and statement combination more than once, then this likely will

tip them off that the repeated sentence is important. To avoid this issue, a Latin Square

Design ensures that each participant will see every possible condition e.g. A Scenario with

S1 (A1), B1, A2, B2, A3, and B3, but they will not see two of the same scenario stimulus

combination.

The fillers here contained scenarios comparing events, individuals, as well as a

variation of other kinds of comparisons including personalities, likes and dislikes, etc. The

possible filler statements were also a mixture of comparatives and declarative statements.

This was done to avoid participants finding patterns of only answering TRUE or FALSE to

events or individual comparisons. Finally, the order in which participants saw stimuli were

randomized.

3.1.1.3 Participants

The participants in this experiment were expectedly quite diverse, since I pulled

participants from a world-wide data-pool. As part of their demographics collections,

participants were asked if they lived in the United States from birth until (at least) age 13,

and if both parents spoke English a majority of the time during those years. If they

answered yes to these questions, we considered them native speakers of English, though

defining who is a native speaker or not can be more flexible depending on the experiment.

Additionally, I collected information on their gender, age, and current residency. Over one

weekend, I collected 96 Turkers, 69F, 25M[8], ages 19-66, with a mean age of 38.56. Each

Turker was paid $3.00 for their time. Participants were removed for the following criteria.

---

[8] Two participants elected not to respond to gender but were included in this analysis.

One, if their accuracies based on our predictions of S1 and S2 sentences was less than 75%. Two, if they indicated that they were not native speakers of English, or three, if they indicated that they were not currently residents of the United States. Each Turker was paid $3.00 for their time, regardless if their data were thrown out or not. However, we believed that this compensation, much greater than the $0.10/min Amazon recommended, would serve as a good incentive to accurately complete this task.

3.1.1.4 Procedure

Each participant from Mechanical Turk was asked to first fill out a preliminary questionnaire that polled their demographic information as discussed in Section 3.1.1.3. Afterwards, they were instructed to click on a link that took them to the Qualtrics survey away from the Mechanical Turk website[9]. While many surveys are done within Mechanical Turk's interface, Qualtrics was used to collect data due to its ease of use and design. From here, participants were randomly presented with forty scenarios one webpage at a time. Each true/false/not sure statement had a forced measure to make sure every question was answered, and they were visually reminded to answer each question if they tried to skip ahead. The order for each survey was randomized using a Qualtrics randomization algorithm, and the surveys themselves were pseudo-randomly distributed, making sure to release an equal number of each type. For 96 participants, that means there were 16 iterations of each type of survey.

After completing their survey, participants were instructed to input their Turk ID into Qualtrics to maintain anonymity as well as to record and verify that the participation

---

[9] HTML code was also used from (Sprouse, 2011), publicly available at
http://www.socsci.uci.edu/~jsprouse/#tools.

was human. Qualtrics did not record any surveys that were incomplete and only completed

surveys were rewarded. To ensure compensation for each Turker, they were given a code

word (*Chocolate*, *Vanilla*, *Strawberry*, *Mint*, *Raspberry*, or *Orange*) that they later entered

into Mechanical Turk's interface to receive their payment.

3.1.2   Results

     After all participants were accounted for and passed through rejection criteria, their data were exported from Qualtrics. Participants were tallied up and raw counts were determined for how many true, false, and not sure answers were given. Microsoft Excel and R (R Core Team, 2013) were used for statistical analyses. Figure 5 shows these raw numbers.
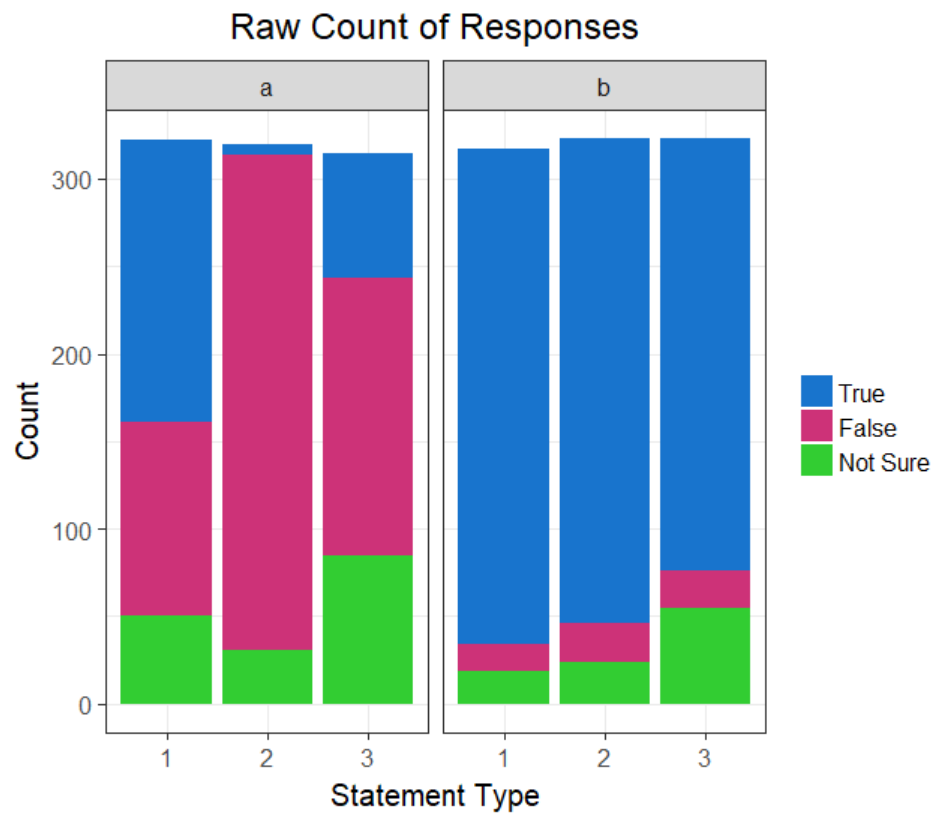
Figure 5 - Raw count of all responses

     In Figure 5, there are two plots strung together, a vs. b scenarios, with the statement type at the bottom. The first observation to make here is that participants are confident in their answers, only answering NOT SURE for %13 percent of their total responses. Following the predictions from Table 1, a couple interesting phenomena become apparent.

First, regarding Statement 1 in Scenario A (A1), it appears that participants were roughly

divided between answering true and false. This suggests that they may not have been

paying attention to the number of people involved. In A2, participants overwhelmingly

select FALSE, following my prediction that participants would correctly identify what party

performed more events. Interestingly here, A3 seems to pattern with A2 as does B3 and B2,

suggesting that readers are paying attention to events when trying to resolve Escher

Sentences. To more clearly see this distinction, not sure data were removed. These data can

be seen below in Figure 6.
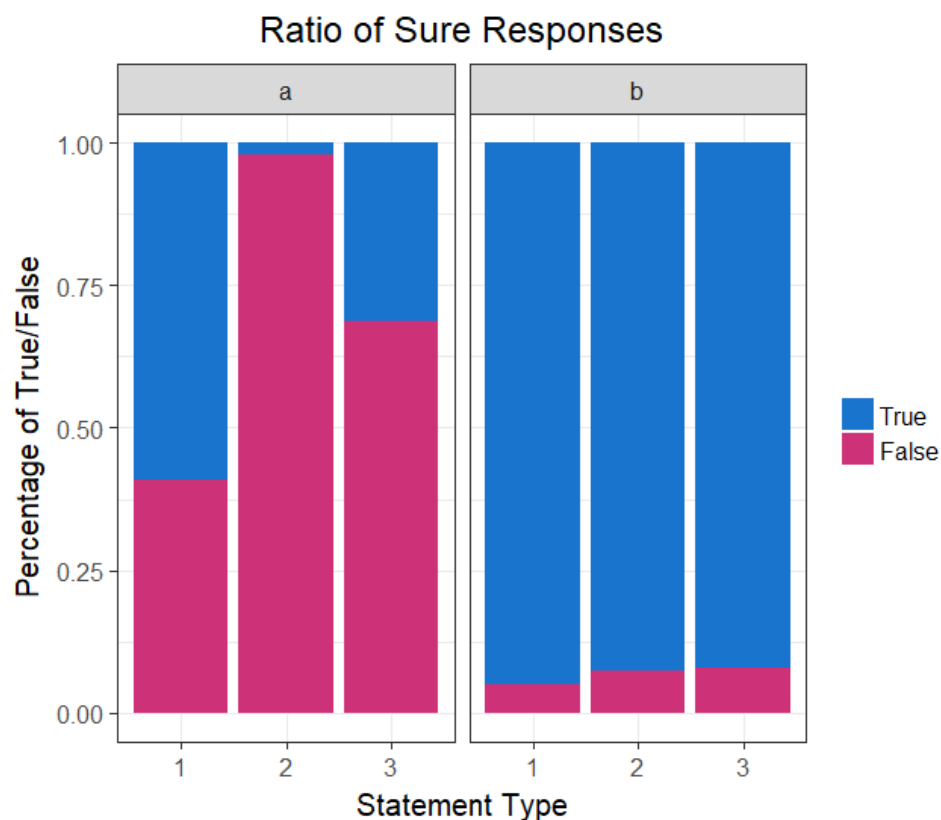
Figure 6 - Ratio of Sure responses.

91

In Figure 6, the not sure data was removed, and the raw count on the y axis has been replaced with a percentage of true and false responses. If we consider each category and take the majority response, A3 is FALSE and B3 is TRUE, which patterns perfectly with participants keeping track of events rather than individuals. Arguably, by only looking at raw data, this is a heuristic approach at best, since these data do not inform on how each individual responded.

To look at these data from an individual level, the data were put through a bivariate correlation analysis. A bivariate correlation means that for every individual, I computed the average values of the proportion of true statements and averaged this across all participants. Then I correlated how each individual rated their S1 vs. their S3s and their S2 vs. their S3s. Table 2 shows these correlation values below.

	AS1 AND AS3	AS2 AND AS3	BS1 AND BS3	BS2 AND BS3
R(95)=	0.511	0.075	0.358	-.116
P =	.001*	.486	.001*	.282

Table 2 – Bivariate correlations per individual of each scenario/statement combination.

The value R indicates a correlation value, where a 0 would mean no correlation, with 1 and -1 indicating strong positive and negative correlations, respectfully. Looking at these data, readers were significantly relating S1 and S3 sentences, meaning they were tracking individuals as opposed to S2s or events. Further discussion of this analysis will be outlined in Section 3.1.3.

3.1.3 Discussion

Given the results of Experiment 1, there appears to be a divide between whether participants associate event interpretations or individual interpretations with Escher Sentences. However, given the lack of not sure responses, these data also suggest that participants have no problem arriving at some interpretation.

There are several points to consider in understanding these conflicting data regarding interpretations. First, participants drawn from this Mechanical Turk pool may not be behaving differently than participants who are drawn locally and in person. While there was a $3.00 compensation for a roughly 20-minute task, it is unclear how this crowdsourcing may affect their responses. Gibson et al. mention that because Amazon pays by the job and not by the hour, there should be some quality control like a comprehension question (Gibson et al., 2011). In this experiment, accuracy assessment was used with a 75% threshold since the task of the participant was solely comprehension questions. Given this, these data may be reliable, but one feature not taken into consideration is how many jobs a Turker does per day. Because a Turker is paid by the job, this would incentivize the Turker to complete as many jobs as they could during the day in order to make enough money for their time to be worth it. This could motivate participants to simply rush through experiments like this. Additionally, given this per job payment, Turkers could have completed several other jobs before completing this experiment, and it is likely that someone who takes this experiment first versus taking twenty before it could drastically affect their attention and their ability to answer questions accurately.

Secondly, while this group is larger than those in typical surveys, there is also a larger age gap. It is the case that speakers will behave differently at age 18 versus age 65,

93

and additionally, because of the anonymous nature of Turking, it is not clear if these ages are accurate. Granted, because each participant is paid anyway, there is no reason to lie about particular demographics (Gibson et al., 2011). However, because this cannot be fact checked, all age groups were included in this analysis.

Regarding the raw data collection, from first glance, including not sure responses, there appeared to be a likeness between A2 and A3 scenarios, as well as B2 and B3, which suggests a correlation between event interpretations and Eschers. However, once the not sure data were removed, these relationships weakened, and in the B scenario cases, each statement was nearly indistinguishable from each other. This could be due to the acceptability of the control sentences. For example, it would be interesting to measure the acceptability of "just me" completions in S1 versus the event comparisons present in S2. Regarding "just me," though, there were several not sure as well as incorrect responses in the A1 category, where a statement like *more trainers lifted weights than just me* would have been elicited. It could be the case that this sentence structure is awkward for participants, which resulted in almost 50% of their responses in the A1 category to be not sure or false.

With respect to the bivariate correlation analysis, this paints a clear picture that participants are relating Escher Sentences more closely with individual interpretation statements. Participants are tasked to make numeral comparisons, and while those numbers can vary, "just me" will always be an easy comparison to make against any number. However, they were largely inaccurate for these statements as well. Thus, it is not clear if participants are relating the interpretation of individuals to Escher Sentence or if they are using some other strategy. To elucidate, imagine the perspective of participants

taking this survey. While they read several instances of comparisons, when they run into an Escher Sentence like *more skiers have fallen down than I have,* this is difficult to interpret. However, if they have just been primed with seeing "just me," which according to the data was also confusing for them, they could relate to Escher Sentence statements in this way, essentially linking the confusing statements together.

All in all, while this data from this survey constitutes a preliminary look into the interpretation of Escher Sentences, it highlights a few important features. One is that Escher Sentences are not straightforward or easy to interpret. This experiment highlights the variability of these interpretation responses, which calls for careful considering in designing future studies to look at how participants understand Escher Sentences. Experiment 5 will return to the issue of what to do to with the problem of interpreting these sentences. Nevertheless, these findings lead us to the next experiment, where I examined how participants accept these sentences rather than how they interpret them.

3.2   Experiment 2 – Untimed Judgment Task

Abstract

Experiment 2 was a behavioral survey that examined the acceptability judgments of Escher Sentences. This experiment divides Escher Sentences into two groups: Strong and Weak Eschers, where Strong Eschers have plural comparative DPs and Weak Eschers have singular comparative DPs. 56 participants rated Strong and Weak Escher Sentences on a scale of 1-7. Results indicate that Strong Eschers are rated significantly less acceptable than controls, and Weak Eschers significantly less acceptable than Strong Eschers.

From Experiment 1, while it appears that participants were arriving at some interpretation, what interpretation they arrived at is still unclear. The next step in understanding the behavioral responses is to look at how participants rate Escher Sentences. According to other researchers, it is the case that repeatability of the predicate and the plurality of the comparative DP shifts the acceptability judgments of Escher Sentences (O'Connor et al., 2012; Wellwood et al., 2009, 2017). In this experiment, I thus set out to not only gather another set of acceptability judgments from participants, but to also see if modulating the plurality of the DP makes a distinct difference between Escher Sentences. Repeatability of the predicate is not included here in order to simplify the design of the stimuli. I divided the Escher Sentences into two groups, Strong Eschers, meaning a stronger illusion or more likely to be rated highly (fooling the participant) and Weak Eschers, meaning a weaker illusion or less likely to be rated highly.

3.2.1   Methods

3.2.1.1 Overview

Acceptability judgment tasks are fundamental to linguistic inquiry. They are one of few tasks available to examine aspects of sentences outside of their own perspective (Schütze & Sprouse, 2014). A linguist is often trained in many languages, theoretical frameworks, and has a lot of experience with analyzing sentences, so simply relying on their own intuitive judgments of sentences may not be sufficient. When running judgment tasks, linguists typically give a series of baseline acceptable and unacceptable sentences. From here, they mix in their experimental sentences and see how these experimental sentences fare against control conditions that would rated high and low. In the case of Escher sentences, there is little research done on how Escher Sentences fit into the spectrum of acceptability. This is the main motivation for this study, to continue to add to the foundation of understanding in how participants view Escher Sentences. Additionally, I will put to the test the difference between Strong and Weak Eschers and see if they are rated distinctly. Later experimentation will also support that these subtypes of Escher Sentences are also neurophysiologically distinct (see Sections 3.3. and 3.4).

For this experiment, I created a written survey that I distributed to students at Michigan State University in 2014. Students were encouraged to take their time to stress an untimed measure of these sentences. In later experiments (see Sections 3.5 and 3.6), I begin to test the distinction between timed and untimed judgments. However, this experiment aimed to provide participants with as much time as they needed without creating more variables for study. This experiment set out to gauge the acceptability

judgment of Escher Sentences, specifically Strong and Weak Eschers, compared with grammatical controls.

3.2.1.2 Stimuli

Experiment 2 utilized sentence triplets to represent each experimental condition. The triplet consisted of a control comparative, a Strong Escher, and a Weak Escher. Between each condition, minimal changes were done to reduce the chance of confounding variables interfering with the acceptability judgments from participants. Thus, each member of the triplet contained practically the same structure save for one target area. Previous research has suggested processing costs at the auxiliary verb following the comparative DP, so this was the exact region to focus on for this experiment (O'Connor et al., 2012). (37) provides an example triplet.

(37)   a. More Brits visited the London Tower than Americans did last year. (Control)

b. More Brits visited the London Tower than the Americans did last year. (Strong)

c. More Brits visited the London Tower than the American did last year. (Weak)

In (37), there is a sentence triplet, labelled Control, Strong, and Weak. The Strong and Weak conditions are the Escher Sentences and crucially differ from the Control condition by simply adding an overt determiner *the* and shifting the plurality of the DP. Wellwood et al. (2009) and O'Connor et al. (2012) both found that the singular comparative DP was markedly worse than the plural, motivating the terms Weak and Strong for the Escher Sentences. Strong means more likely to elicit an illusory effect i.e. be assigned a higher acceptability rating, and weak means less likely to elicit an illusory effect.

Furthermore, previous research has also shown that if the matrix clause contains a repeatable predicate, then this increases the likelihood of an Escher Sentence being rated more highly (O'Connor et al., 2012; Wellwood et al., 2009). Because one manipulation for plurality in the DP has been done in (37), all sentences in this experiment have repeatable predicates throughout to minimize variable manipulation. These triplets were evenly distributed amongst four different surveys, A, B, C, and D using a Latin Square Design. Each survey contained a 1:1 ratio of fillers to experimental sentences, with 30 experimental and 30 fillers, totaling 60 judgment responses. This, like all experiments in this project, was a within-subjects design, meaning each participant saw every condition.

Fillers consisted of a mixture of both acceptable fillers (good fillers) and unacceptable fillers (bad fillers). This was done to establish a floor and ceiling for how to rate these sentences. Additionally, fillers contained a mixture of comparative and declarative sentences to prevent participants from noticing the experimental cases contained only comparative sentences.

For all acceptability tasks in this project, a Likert Scale from 1 to 7 was utilized (Likert, 1932). Participants were instructed that they were rating acceptability, where acceptability meant how they personally would accept that sentence if they heard someone say it e.g. 1 = not an acceptable sentence of English and 7 = completely acceptable sentence of English. The term grammatical was avoided in the instructions for participants, not only because this study focuses on the acceptability of the sentence, but that most naïve participants have preconceived notions of what grammatical means that could potentially affect their results.

3.2.1.3 Participants

There were 56 participants, 35F, 22M, ages 18-35, with a mean age of 20.9 years old. Participants were recruited from MSU linguistics classes and were offered extra-credit for their participation in this experiment. I was not the instructor for any of these students, removing any instructor bias that students may have for me as an experimenter.

3.2.1.4 Procedure

Because this was a written survey, each version was printed and color-coded for ease of distribution. On the survey, the sixty sentences were randomly scrambled using a random number set generated from www.random.org. Each sentence had a box for participants to leave their judgment using a Likert Scale of 1 to 7. No personal information was recorded on the survey, but students were asked to report age and gender near at the beginning of the experiment.

I asked the instructor of each class I used at least two weeks beforehand to leave 20-30 minute available near the end of class to complete this survey. When I arrived, each student was given a consent form from our lab to ensure anonymity of their data. Students were then instructed to read each sentence carefully and rate the acceptability of every sentence, where a 1 would equate to an acceptable sentence of English and a 7 would equate to a perfectly acceptable sentence of English. After students were done, they were instructed to put their pencils down and to wait for the rest of the class to finish. This was done to ensure that there were no interruptions in concentration from students leaving randomly throughout the process. Each survey was hand-tallied and recorded using Microsoft Excel.

3.2.2   Results

After the responses were recorded, the mean acceptability judgments for each sentence type across individuals was calculated. Figure 7 below shows the mean responses of each sentence type across participants.
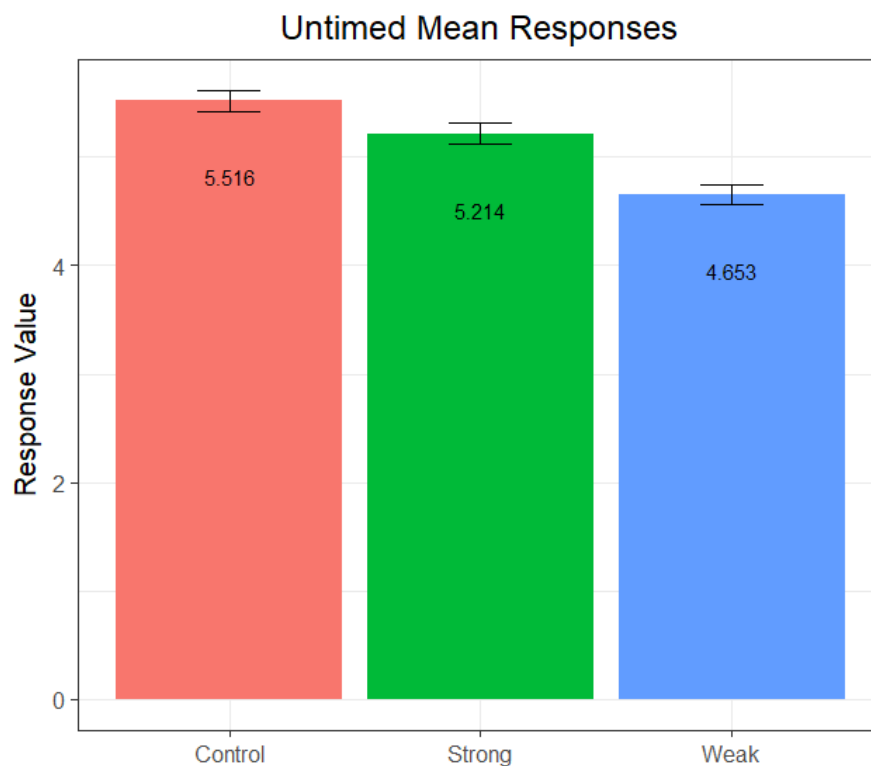
Figure 7 – Untimed mean responses for Control, Strong, and Weak Escher sentences. Error bars represent the standard error.

In Figure 7, the mean value is indicated within each bar, and each bar has an error bar at the top indicating standard error. Standard error is the standard deviation, or standardized units away from the mean, of the sample size, here 56 students. Error bars are handy, because if a neighboring bar is under it, one can assume that the difference between conditions is statistically different. Statistically, a paired student's t-test indicates

a significant difference between Control and Strong conditions where [t(55)=3.199, p=.002*]. A p-value less than .05 is typically considered statistically significant, which will be the standard used throughout this project: an asterisk will indicate significance. To elucidate, a p-value of less than .05 means that there is less than a 5% chance that the difference found was just random chance, or, there is more than 95% confident these two conditions are different. A paired student's t-test also indicates significance between Strong and Weak Escher conditions, where [t(55)=4.749 ,p<.001*].

The mean responses of these judgments suggest that Control, Strong, and Weak Escher conditions are different. However, it is also important to look at the distribution of how participants respond to these experimental sentences. By looking at the distributions of how participants respond, we can determine the overall shape of the data and see if there are any interesting patterns in how they responded. These distributions are shown in a series of histograms in Figure 8.

Untimed Distribution
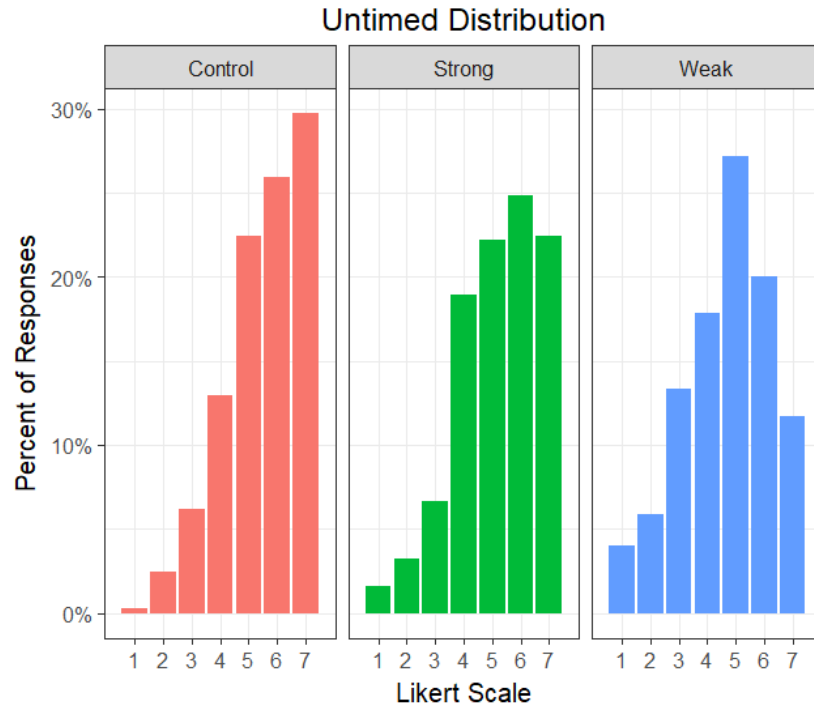
Figure 8 – Untimed distributions of Control, Strong, and Weak Escher sentences

In Figure 8, each condition receives its own plot, and on the x-axis, each Likert-Scale value is shown from 1 to 7. To ascertain a better understanding of the distribution of responses, the percentage of each Likert value is shown rather than a raw count. In the Control condition, a majority of responses are rated 5-7, whereas the majority of response for Strong and Weak are more towards the middle of the scale. These visually highlight a degradation of acceptability in Escher Sentences as well as an inherent variability of ratings, which is a unique pattern for Escher Sentences. Compare, for example, the good and bad fillers in this experiment and their overall shape, shown below in Figure 9.
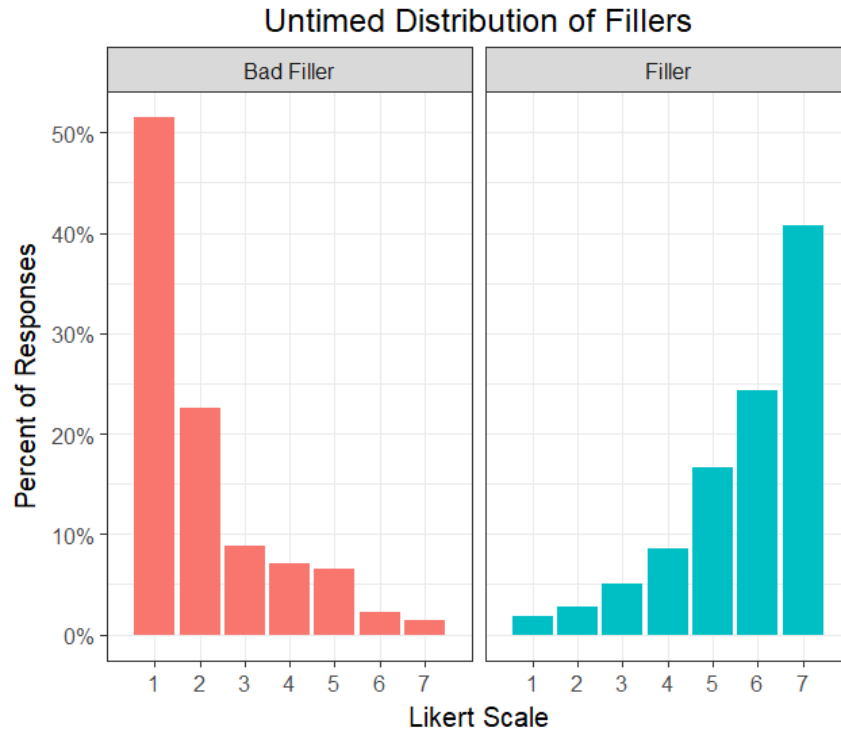
Figure 9 - Distribution of good fillers vs. bad fillers with proportion values.

These distributions are more standard for what a linguist would expect for either a good sentence (filler) or a bad sentence (bad filler). While looking at these judgments provides an interesting insight into participant's judgments of acceptability, there are a few flaws to immediately address. First, when using parametric statistical tests like a paired student's t-test, the assumption is that the errors follow a normal distribution (Kim, 2015). Likert Scale judgments like these, however, do not follow a normal distribution because Likert Scales are bounded by 1 being the lowest and 7 being the highest. Moreover, it is not the case that participants use the Likert Scale the same way: some people may only be using 4-7 with four as their lowest, others may use the whole range or some subset of these numbers. Thus, simply using parametric t-tests with biased-scale Likert Scale judgments is inherently flawed, so the best solution here is to take the data and shift it towards a

continuous scale for each participant. This can be done by using a z-score transformation on each subject (Schütze & Sprouse, 2014).

First, for each participant, their average rating for all their judgments, including fillers, was calculated. Then, each participant's average judgment value was subtracted from each of their individual judgments. This difference was then divided by the standard deviation of the participant's overall average. By doing this, each judgment is no longer on a scale of 1-7, but instead is a standardized version of each individual's score, represented as standard deviations from that participant's mean (Schütze & Sprouse, 2014). The parametric t-tests used before this transformation are not as reliable as ones performed on z-score transformed data. Therefore, while I have shown how Likert Scale data can be represented without any transformations, for the remainder of this dissertation project, I will only be using z-score transformed data to more reliably represent the judgments of Escher Sentences.

Figure 10 shows the mean responses for each condition now with a z-transformation on the judgment data from the participants.
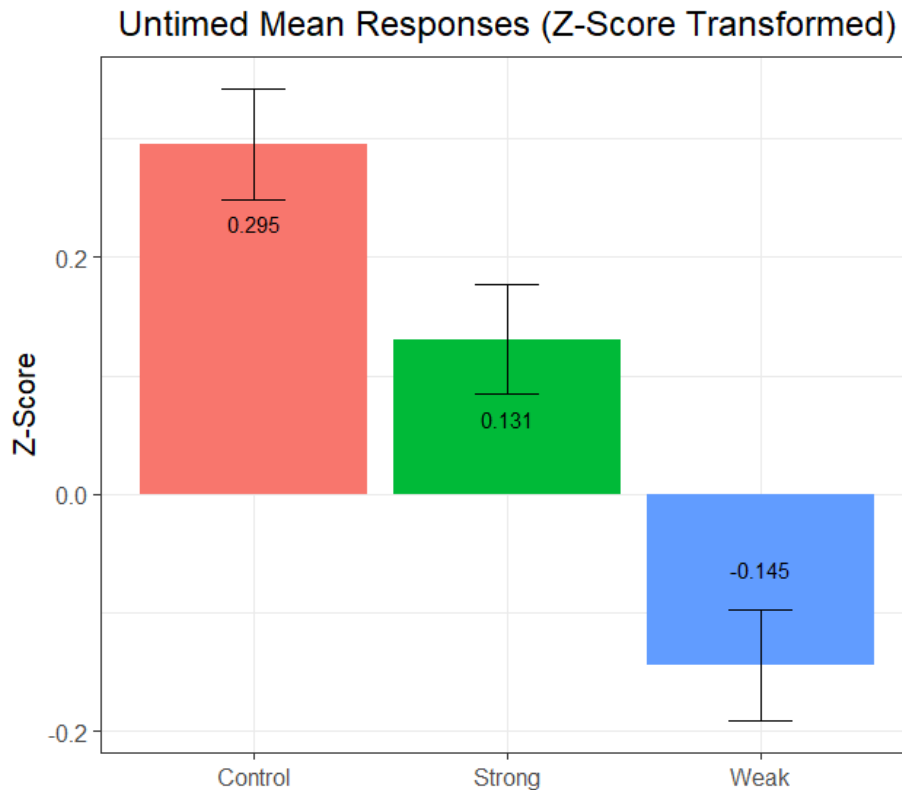
Figure 10 – Untimed mean responses of Control, Strong, and Weak Eschers with a z-score transformation.

Notice here that the relationships between each group remain, namely that Controls are rated significantly higher than Strong Eschers, and Strong Eschers are rated significantly higher than Weak Eschers. To note, because z-scores operate based on standard deviations away from a mean, there will now be negative values, even though participants are still only using values from 1 to 7. To clarify, positive values would indicate that ratings were above the mean rating, which always equals 0 in these transformations, and negative values indicate ratings below this mean. With z-transformations, each participant has a standardized scale according to their own rating system, eliminating any bias that could exist with Likert Scale ratings across participants.

Table contains the paired student t-tests between Control, Strong, and Weak Escher conditions.

Comparison	Non-Transformed	Z-Score Transformed
Control vs. Strong	t(55)=3.199, p=.002*	t(55)=3.424, p=.001*
Strong vs. Weak	t(55)=4.749 ,p=.00001*	t(55)=4.837 ,p=.00001*

Table 3 – P-values of z-score transformed data compared with earlier non-transformed t-tests.

For this set of data in Experiment 2, comparing the values in Table 3 with the previously calculated differences shows very little change. Nevertheless, z-transformations are more reliable when using parametric statistical tests.

With respect to the distribution of responses for these data, a z-transformation affects them quite differently. For z-transformed judgment data, using a histogram poses a great challenge, as z-scores no longer fit nicely into a categorical rating system. Instead, a point-graph with a best fit line will be the best representation of distributions for these sentences, demonstrated in Figure 11 for the distribution of experimental conditions and Figure 12 for the distribution of fillers.
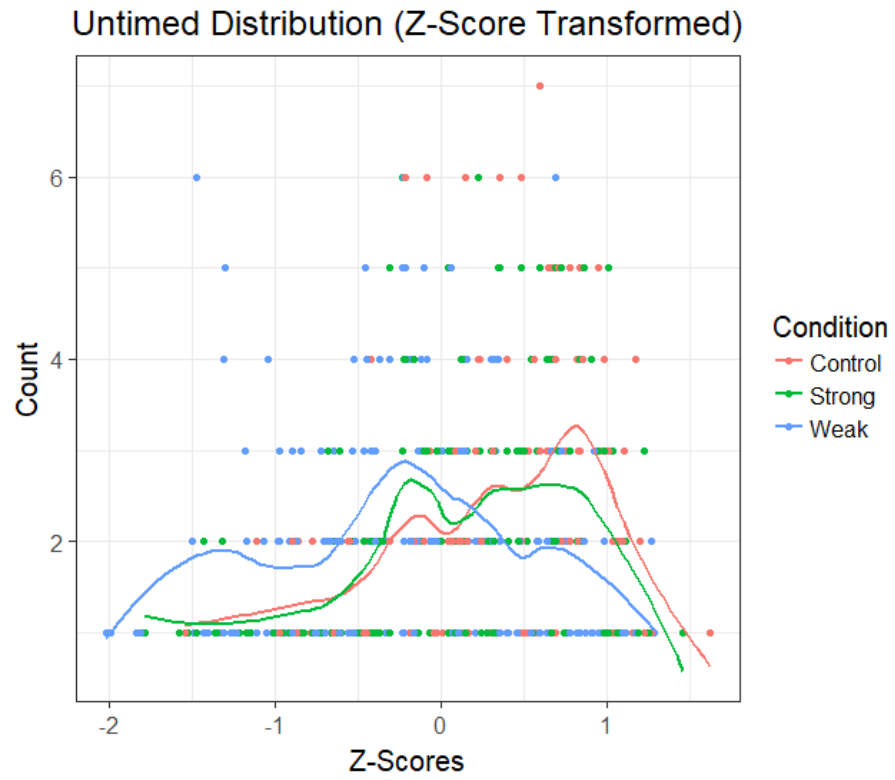
Figure 11 – Distributions of Control, Strong, and Weak Eschers with a z-score transformation.

Untimed Distribution of Fillers (Z-Score Transformed)
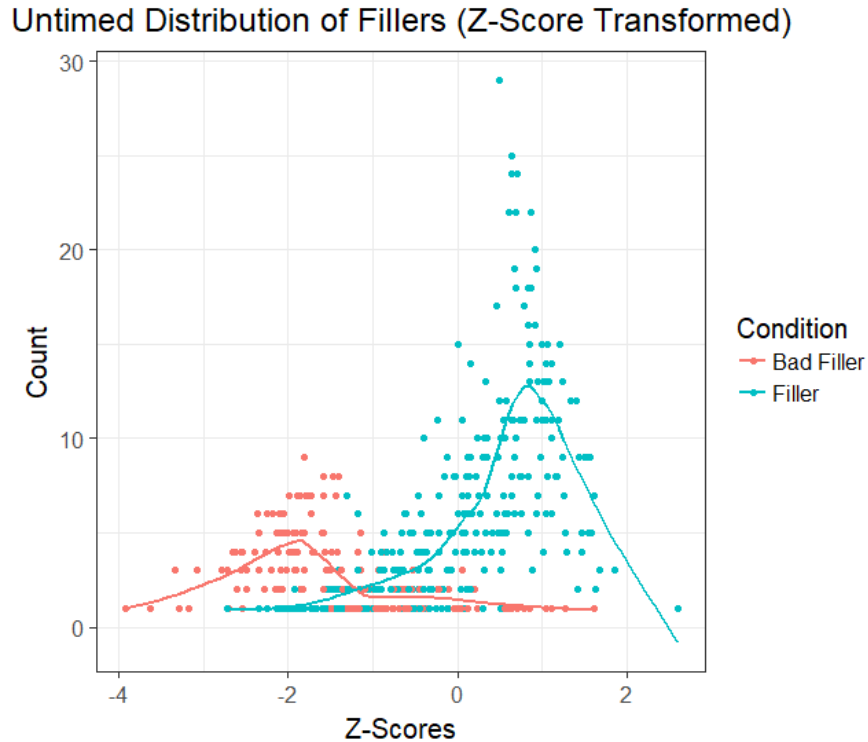
Figure 12 – Distribution of good vs. bad fillers with a z-score transformation.

In these figures, each z-score was plotted for each participant, with counts of each individual value shown on the y-axis. A best fit line was calculated for each condition using Loess Regression method, where Loess is short for local regression (Cleveland, 1979). Visually, the Control and Strong Escher conditions veer towards the right of the z-score pattern, meaning they are rated roughly one standard deviation on average above the mean, whereas Weak Eschers are a roughly .5 standard deviations below the mean. Visually, these graphs are quite distinct from the histograms earlier and provide a clearer picture of how participants were rating these sentences. To determine differences between these conditions, a one-way ANOVA, or analysis of variance, revealed a significant effect of condition [$F_{(2,110)} = 22.043$, $p < .001*$]. For fillers, a one-way ANOVA revealed an effect for condition [$F_{(1,55)} = 354.033$, $p < .001*$]. These results come at little surprise given the

visual distinctions between filler conditions and experimental conditions. Furthermore, when looking at fillers, there is a clear pattern of what acceptable and unacceptable sentences (good and bad fillers, respectively) look like. For the Escher Sentences, their shape is quite different: rather than leaning towards one end of a standard bell-curve, they lean more towards the mean or center. This is indicative of sentences that are variable rather than clearly acceptable or unacceptable.

3.2.3   Discussion

This experiment examined the acceptability judgments of Strong and Weak Eschers versus Control comparatives. In the observed data, both Strong and Weak Eschers were rated significantly lower than the Control condition. Strong Eschers were significantly more acceptable than Weak Eschers, but they were significantly less acceptable than the Control comparatives. This experiment also looked at the distributions of Escher Sentences, Controls, and fillers. Visually, Escher Sentences appear to be closer to a mean rating, indicating that participants varied in how they treat an ungrammatical Escher Sentence.

The variability of judgment responses is of great interest to this project, and this experiment highlights this variability. If it were the case that participants were consistent in their judgments, the distribution of Escher Sentences should resemble other ungrammatical sentences, like the bad fillers in this experiment. Alternatively, if participants were always fooled by Escher Sentences, then the Escher Sentence distribution should match that of grammatical sentences, like the good fillers in this experiment. The elicited judgments of Escher Sentences beg for a reexamination of the general and anecdotal claim that participants, at first glance, are accepting these sentences. In an untimed setting such as this one, one could predict that the illusion should fall apart if a participant is given more time to think about these sentences. The data from this experiment, however, suggest that while participants are clearly recognizing that Escher Sentences are less acceptable than grammatical Controls, participants also demonstrate that they believe Escher Sentences are not as bad as other ungrammatical sentences like the bad fillers. While this experiment does not directly test the claims made by AET, I suggest that the reason for this variability stems from the matrix event-individual

111

ambiguity and unlicensed ellipsis. Further exploration of this topic occurs in Sections 3.5 and 3.6.

There are a few remaining concerns for these data. First, while the Control condition was shown to be significantly different than both Escher conditions, the distribution of Control sentences and Strong Eschers were rather similar. This could suggest that Strong Eschers and Controls have the potential to have no difference between them. If anecdotal observation of Escher Sentences is true, then this may be due to participants spending more time looking over Escher Sentences and attenuating their illusory effect. One way to test this will be to present these sentences in a timed setting for participants to see if limiting their time encountering these sentences will make participants more likely to equate Control comparatives and Strong Eschers. This line of inquiry is conducted in Experiment 5 in Section 3.5.

Furthermore, the analysis of why Strong Eschers are more likely to elicit an Escher behavior, meaning to "fool" the participant into thinking the Escher Sentence is acceptable, potentially lies with the Strong Escher's closeness to a grammatical comparative control. In the case of the control, there is a bare DP as the comparative DP where in the Strong Escher the DP is overt and plural and in the Weak Escher it is overt and singular. By deviations away from the control, bare DP, Weak Eschers are two deviations away (singular, overt), and Strong Eschers are only one deviation (overt). If it is the case that the parser cares about features of an error, then this is possibly a syntactic and semantic reason why Strong and Weak Eschers behave differently.

Nevertheless, the distribution of the Control sentences in this experiment could also simply be due to a lack of acceptable controls. In stimulus design, sometimes a sentence

generated could have awkward wording or structure, and participants can be sensitive to this. One way to avoid a problem like this would be to run a pretest on a separate group of participants to determine how acceptable the Control sentences are. Any sentences that are found to be below some threshold could be excluded.

Another issue that could have potentially interfered with this experiment was the usage of the word *acceptable* when instructing participants to rate sentences. After the acceptability judgment experiment, participants reported confusion as to what *acceptable* meant, often hyper-correcting their own definition to mean *grammatical*. To avoid any issues that this may cause for these data in this experiment, or for any further data, I frame each acceptability task as *natural*, meaning a 1 would be *not a natural sentence of English* and a 7 would indicate *a completely natural sentence of English*. While I have no experimental data to back up my claim that *natural* is a better term to use, for each future experiment, it did transpire that there was significantly less reported confusion on what the task entailed. This would be an interesting method project for a future experimenter to explore.

3.3   Experiment 3 – EEG Strong vs. Weak Eschers

Abstract

Experiment 3 used EEG to compare two kinds of Escher Sentences, Strong and Weak Eschers. 17 participants encountered 180 sentences under a rapid serial visual presentation (RSVP) and were asked to rate each sentence on a Likert Scale of 1-7. Behavioral results indicated a replication of a significant difference in acceptability between Strong and Weak Escher Sentences. Electrophysiological suggest an early effect in the 75-115ms post stimulus window. These results suggest early recognition of an unexpected and unlicensed ellipsis.

Experiments 1 and 2 have established two important facts regarding Escher Sentences. One, while it is unclear what interpretation participants arrive at, they clearly are interpreting something with respect to Escher Sentences. Two, there is great variability in how participants accept these sentences, though they are less acceptable than grammatical controls. Experiment 3 asks how the brain responds to the difference between Strong and Weak Escher Sentences. Furthermore, Experiments 3 and 4 seek to determine whether it is true that the parser actually ignores the problem of the Escher Sentence or if it is able to detect an issue (see Ferreira & Patson, 2007; Phillips et al., 2011). This experiment examines what is happening in real time when people encounter Escher Sentences.

3.3.1 Methods

3.3.1.1 Overview

This experiment pioneers the field's understanding of Escher Sentences, and the comparison of interest here is between the Strong and Weak Escher Sentence. While this experiment has no grammatical comparative condition, it is not necessary to determine if any difference exists between Strong and Weak Eschers given Experiment 2.

Due to a lack of experimental findings with respect to event related potentials (ERPs) and Escher Sentences, it is difficult to predict what kind of ERP to expect. However, using EEG should highlight potential costs incurred from processing these sentences. These costs could manifest as a repair process either semantically or syntactically, which may elicit either an N400 or P600 (beim Graben, Gerth, & Vasishth, 2008; E. F. Lau et al., 2008; Neville, Nicol, Barss, Forster, & Garrett, 1991). However, because these sentences do not behave like typical ungrammatical sentences, it is unclear if there will be any repair cost associated with Escher Sentences. One expectation is that there will be recognition of a syntactic error, which may manifest, for example, as an ELAN, which would be indicative of an early recognition of some syntactic problem (Friederici, 2002; E. F. Lau et al., 2008; Steinhauer & Drury, 2012). Alternatively, another possible early recognition of change could be from a visual N100, typically associated with changes in attention and unexpected stimuli (Luck et al., 1994; Mangun & Hillyard, 1991; Vogel & Luck, 2000).

3.3.1.2 Stimuli

The first step in crafting this experiment was to construct a series of Escher Sentences that mimicked those tested in previous experiments. 60 doublets of Strong and Weak Eschers were created, counterbalanced using a Latin Square Design between two

experimental groups, A and B, such that each group contained an equal number of experimental sentences. This was a within-subjects design, with a 1:1 ratio of filler to experimental sentences. Fillers differed from experimental sentences anywhere from three to seven words in length, containing both declarative and comparative elements. (38) below shows an example doublet.

(38)   a. More knights fought dragons than the peasants did because of the massive fires engulfing the countryside. (Strong)
       b. More knights fought dragons than the peasant did because of the massive fires engulfing the countryside. (Weak)

In (38), the Strong and Weak Eschers were crafted to have similar structures from the previous experiments in this dissertation project. The Strong Eschers had a plural DP in the comparative clause and the Weak Eschers had a singular DP in the comparative clause. Additionally, end of sentence corrections were applied to the Escher Sentences to avoid sentence wrap-up effects that participants have when arriving at the end of a sentence (Hirotani, Frazier, & Rayner, 2006; Rayner, Kambe, & Duffy, 2000). Since the target area for this EEG experiment would be at the end of the sentence, this necessitated end of sentence corrections ranging anywhere from four to ten words long. Importantly, any content after the target area did not reference the comparative DP or the matrix DP to avoid any kind of long-distance grammatical relationships that could affect processing (Clifton & Frazier, 1989). In other words, having the end of sentence correction refer back to the original

subject could incur processing costs associated with maintaining in memory the subject from several words prior.

3.3.1.3 Participants

22 participants[10] were recruited for this experiment, 10M, 11F, ages 18-24, with a mean age of 20.6. Participants were recruited from MSU linguistics classrooms and were given extra credit for their time. I was not the instructor for any of these students to avoid instructor-experimenter bias.

3.3.1.4 Procedure

Participants were first welcomed into the MSU Neurolinguistics Lab and asked to read over and sign a consent form. They were then instructed on the entire process which highlighted a few educational aspects of EEG as well as safety and general protocol. Each participant sat comfortably in a sound-attenuated room separate from the experimenter. After being individually fitted for a cap, experimenters began inserting gel into each electrode site. Upon completion, participants were shown their own EEG waveforms live using ASA software (Advanced Neuro Technology) to aid in the educational process as well as to comfort them in knowing that there is no risk in EEG recording. They were then instructed to pay attention to the computer screen in front of them, which directed them to use a Cedrus Button Box (www.cedrus.com) containing seven buttons numerically labeled 1-7. 120 sentences were presented word-by-word using a rapid serial visual presentation (RSVP) paradigm which included 30 strong Eschers, 30 weak Eschers, and 60 fillers. Each sentence began with a fixation cross at the center of the screen for 1000 ms followed by a word-by-word presentation in the center.

---

[10] Demographic information for one participant is missing due to hard drive failure.

Each word had a 600ms stimulus onset asynchrony (SOA), with presentation for 400 ms and an inter-stimulus interval (ISI) of 200 ms. These times were chosen after a pretest had confirmed that this was a comfortable pace for reading. Upon completion of each sentence presentation including fillers, the computer presented a question that stated, "how acceptable was that sentence?" and paused until the participant responded. A task is quite common during EEG and RSVP studies in order to keep the participant focused and on track. Additionally, questions were asked after each sentence to gather participant behavioral data as well as avoiding any kind of pattern effect that might occur if participants were to only see questions after experimental questions.

Participants were instructed to press 1-7 on the button box that corresponded with the acceptability of each sentence, where 1 was "not a natural sentence of English" to 7 which was "a natural sentence of English." For these answers, natural was explained to mean a sentence that felt completely natural or ordinary when talking with a friend. After responding, a fixation cross appeared again, and the next trial began. The same procedure was repeated until responses to all 120 sentences were collected. The session lasted roughly 30-40 minutes per person. The order of sentences for each participant was presented randomly using a randomization technique through Super Lab (www.cedrus.com).

Data were recorded using a 32 sintered Ag/AgCl electrode cap (GND WaveGuard 64 Electrode cap; Advanced Neuro Technology BV., Enschede, The Netherlands). The data were amplified using a Full-band EEG DC Amplifier (Advanced Neuro Technology) with a sampling rate of 256 Hz. Both right and left mastoids were used as a reference, and all signals were captured continuously.

Behavioral and EEG data were gathered for all participants; EEG data were exported, meaning that the analog signal collected by the EEG software was converted to a digital signal using ASA. Behavioral data were exported from the button-presses in SuperLab to a large collection of .txt files. From here, the EEG data were processed into large .csv documents containing time stamps and voltage values using MATLAB (Mathworks, 2018). Statistical analyses was conducted on both behavioral and EEG data using R (R Core Team, 2013).

Regarding the processing for behavioral data, the Likert-Scale ratings were converted to z-scores. For EEG data, the average ERPs were calculated across participants for both conditions, Strong and Weak. The measured epoch was marked 200 ms before presentation of the auxiliary verb after the comparative phrase subject DP and extended 1000ms post-presentation for a total window of 1200 ms. Six ROIs were constructed, defined as— Anterior Left (Fp1, F7, F3, FC5), Anterior Right (Fp2, F4, F8, FC6), Anterior Central (Fz, Fpz, FC1, FC2), Posterior Left (CP5, P7, P3, O1), Posterior Right (CP6, P4, P8, O2), and Posterior Central (CP1, CP2, Pz, POz, Oz,). These ROI were constructed partially on an EEG study done by Whelpton and company (Whelpton et al., 2014). These six ROI were considered as factors for ANOVA Anteriority (Anterior, Posterior) and Laterality (Right, Central, Left).

Contaminated epochs were rejected and double-checked via visual inspection. After visual inspection, MATLAB initiated a moving window artifact rejection that checked every 20ms for voltages beyond a threshold set at +/- 150 µV, resulting in approximately 13% trial rejection. Subjects that were two standard deviations or more away from the mean rejection rate were thrown out; this yielded the rejection of three participants. The ERP

waveforms were quantified by mean amplitude measures in relation to a 200 to 100 ms

pre-stimulus baseline. ROIs and conditions were subjected to repeated measures ANOVAs

using R (R Core Team, 2013).

3.3.2   Results

3.3.2.1 Behavioral

After performing a z-score transformation, the average responses for each condition across participants was calculated. Figure 13 shows the mean responses between Strong and Weak Escher Sentences.
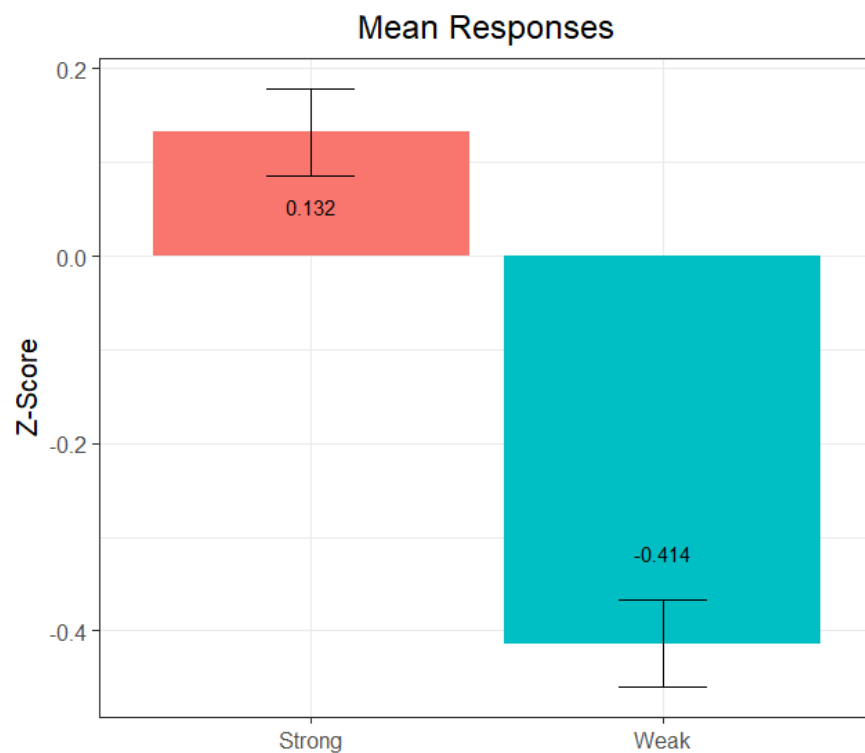
Figure 13 – Mean responses to Strong vs. Weak Escher sentences.

These behavioral data replicate the findings in Experiment 2, where a paired student's t-test indicates significance between conditions [t(16) = 5.879, p=.016][11]. However, there appears to be a much stronger difference between conditions here than in Experiment 2. Next, the distributions of responses were calculated and plotted in Figure 14.

---

[11] This t-value reflects data loss of participant's behavioral data due to software and hard drive errors; this loss of subjects should not greatly affect the trend of these behavioral data.
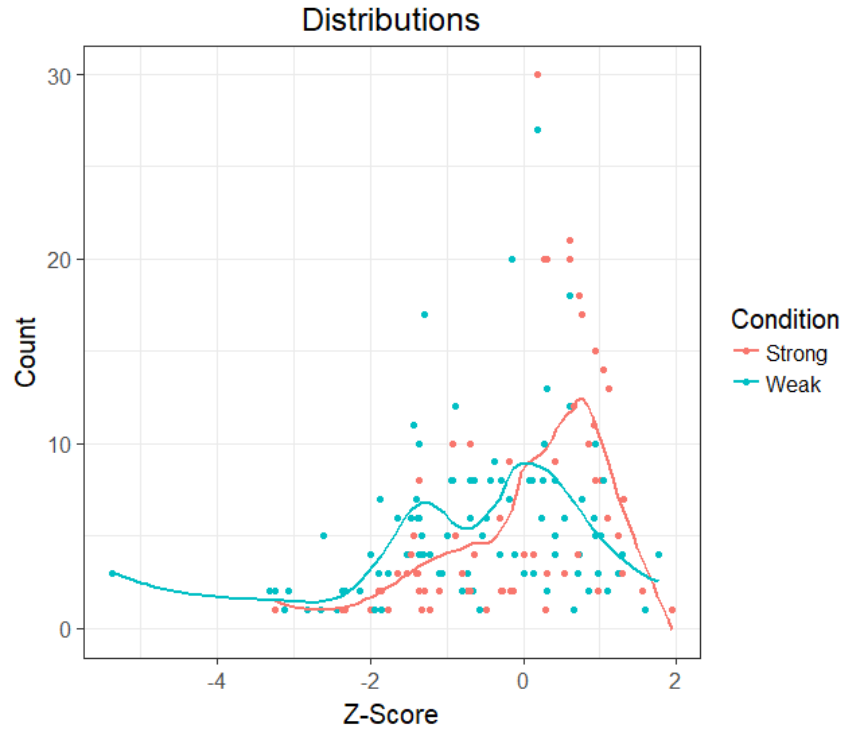
Figure 14 – Distributions of Strong vs. Weak Escher conditions.

In Figure 14, distributions of Strong and Weak Eschers appear distinct from one another. A one-way ANOVA reveals a significant difference between these groups, though not a strong one [$F(1,16) = 4.444, p=.05$]. To follow procedure from previous experiments, the distribution of fillers was also conducted for this experiment, shown in Figure 15. This experiment lacked bad fillers, which will be touched upon in Section 3.3.4.
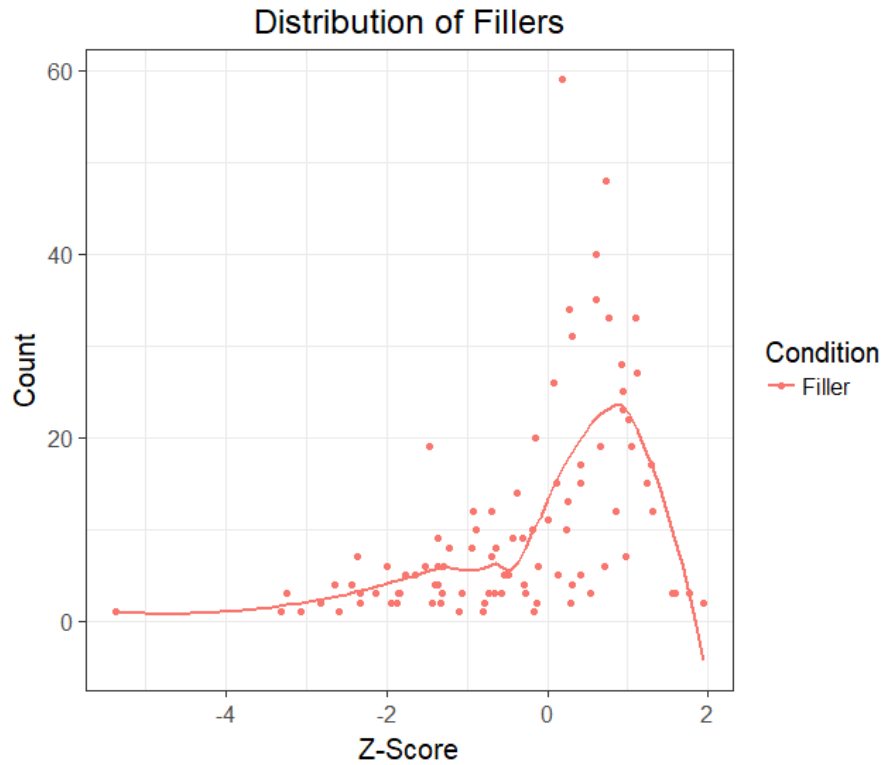
Figure 15 – Distribution of filler stimuli.

3.3.2.2 EEG

Regarding the EEG data, the mean amplitudes for each ROI and each condition across participants were calculated and plotted in Figure 16 below.

Figure 16 - Grand average of Strong vs. Weak Escher conditions.

Figure 16 shows an example of six grand averages, meaning the average voltage of a specific time window (-200 to 1000ms) at a specific area (Anterior Left, Anterior Central, Anterior Right, Posterior Left, Posterior Right, and Posterior Central). The 0ms mark is the point at which the word *did* flashed on the screen for participants. To elaborate, *did* across both Strong and Weak Escher conditions was the consistent target verb after the comparative DP e.g. *than the squire/squires did.* The data before 0ms is baselined data, or data that has been subtracted across the entire dataset as is standard in EEG practice; this

is done to remove a possibility of effects before the window of interest interfering and possibly encouraging a false positive effect. So, across all participants, these six grand averages show their brain activity on average at the point where they encounter the target area *did.* In EEG work, our goal is to determine if there is a difference between conditions, and the highlighted green portions in the Posterior Left, Posterior Central and Posterior Right ROIs indicate areas of significance where Strong and Weak are different. Though visually there appear to be other distinct differences in the data, these are not statistically significant. The highlighted portions represent a window period of 75-115ms. At these regions, an ANOVA was performed using the factors Anteriority (2: Anterior, Posterior), Laterality (3: Central, Left, Right) and Condition (2: Strong, Weak) for a 3x2x2 design. ANOVA is used as a parametric test instead of paired t-tests due to making comparisons between more than two groups.

There were main effects for Anteriority $F(1,21) = 5.786$, p=.025, Laterality $F(2,42) = 6.749$, p=.003, and Condition $F(1,21) = 4.971$, p=.036. For interactions, Laterality x Condition $F(2,42) = 3.464$, p=.041 was significant, but Anteriority x Condition $F(1,21) = 2.298$, p=.144 and Anteriority x Laterality x Condition $F(2,42) = 0.359$, p=.700 were not significant.

Because only partial interactions were observed, the contrasts on which post-hoc analyses can be conducted are limited. In this case, the interaction of Laterality X Condition allows us to look at entire regions rather than one of six specified ROIs. The early effect here is spread across the entire posterior region, where posterior left $F(1,21) = 10.481$, p= 0.004*, posterior right $F(1,21) = 6.061$, p= 0.023*, and posterior central $F(1,25) = 11.211$, p=.003*.

To help to visualize the early effects presented in Figure 16, Figures 17 and 18 show scalp plots of these data.
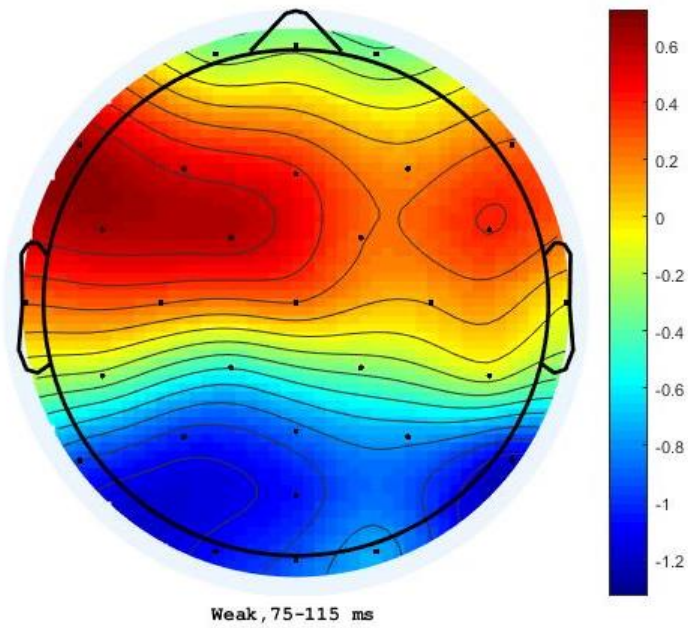
Figure 17 – Scalp plot of the Weak Escher condition with mean amplitude between 75-115 ms displayed.
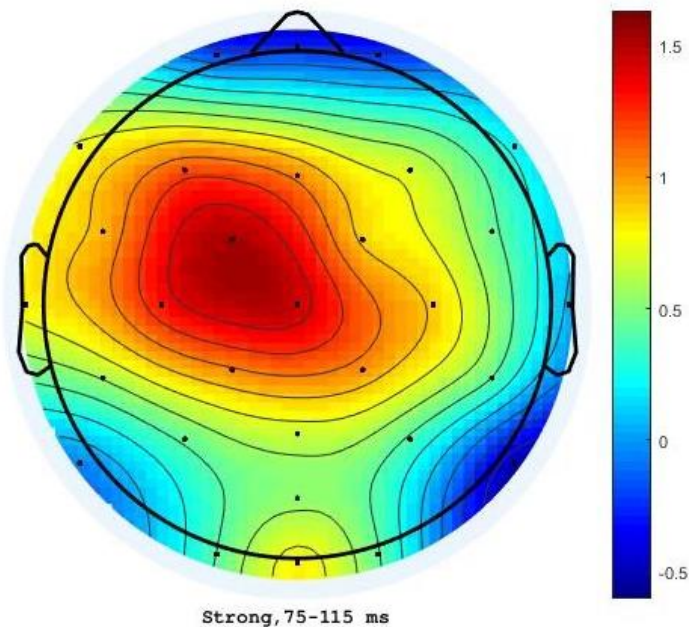
Figure 18 - Scalp plot of the Strong Escher condition with mean amplitude between 75-115 ms displayed.

Figures 17 and 18 show average voltages at the 75-115 most stimulus window for Weak Escher and Strong Escher conditions, respectfully. These scalp plots show the mean voltages at the specified time windows across the entire scalp. Of importance here is to note the posterior regions of the scalp, where in the Weak Escher condition (Figure 17), there is a negativity spread across the posterior region indicated by the blue color, and in the Strong Escher condition (Figure 18), this same region is far more positive, indicated by yellowish green colors.

3.3.3 Discussion

Experiment 3 offers the first findings for Escher Sentences and grammatical illusions with an ERP paradigm. Importantly, the early deflection in the waveform across the posterior region suggests activation of the visual system with respect to some change in attention or unexpected stimulus, most likely an N100 (Luck et al., 1994; Mangun & Hillyard, 1991; Vogel & Luck, 2000). While it is unclear if the deflection detected between Strong and Weak Eschers is positive or negative, this is early evidence to suggest that the parser picks up on some anomaly early on. Since we cannot conclude whether this effect is negative or positive because of a lack of a grammatical comparison, Experiment 4 will attempt to replicate these findings and to determine what type of deflection this may be. One potential argument here could be made that this difference found between Strong and Weak Eschers is due to the different morphological forms of the comparative DP e.g. *the Americans* versus *the American_,* and that the visual C1 complex is responding to this difference. However, this difference is occurring 600 ms before the 0 marker of the data here, which was the always consistent *did*, which likely negates any kind of effect due to earlier data. Figure 19 below shows 700ms before the 0 marker. The same rejection and filtering criteria were applied to these exported data.

Grand Average -700 to 0 ms
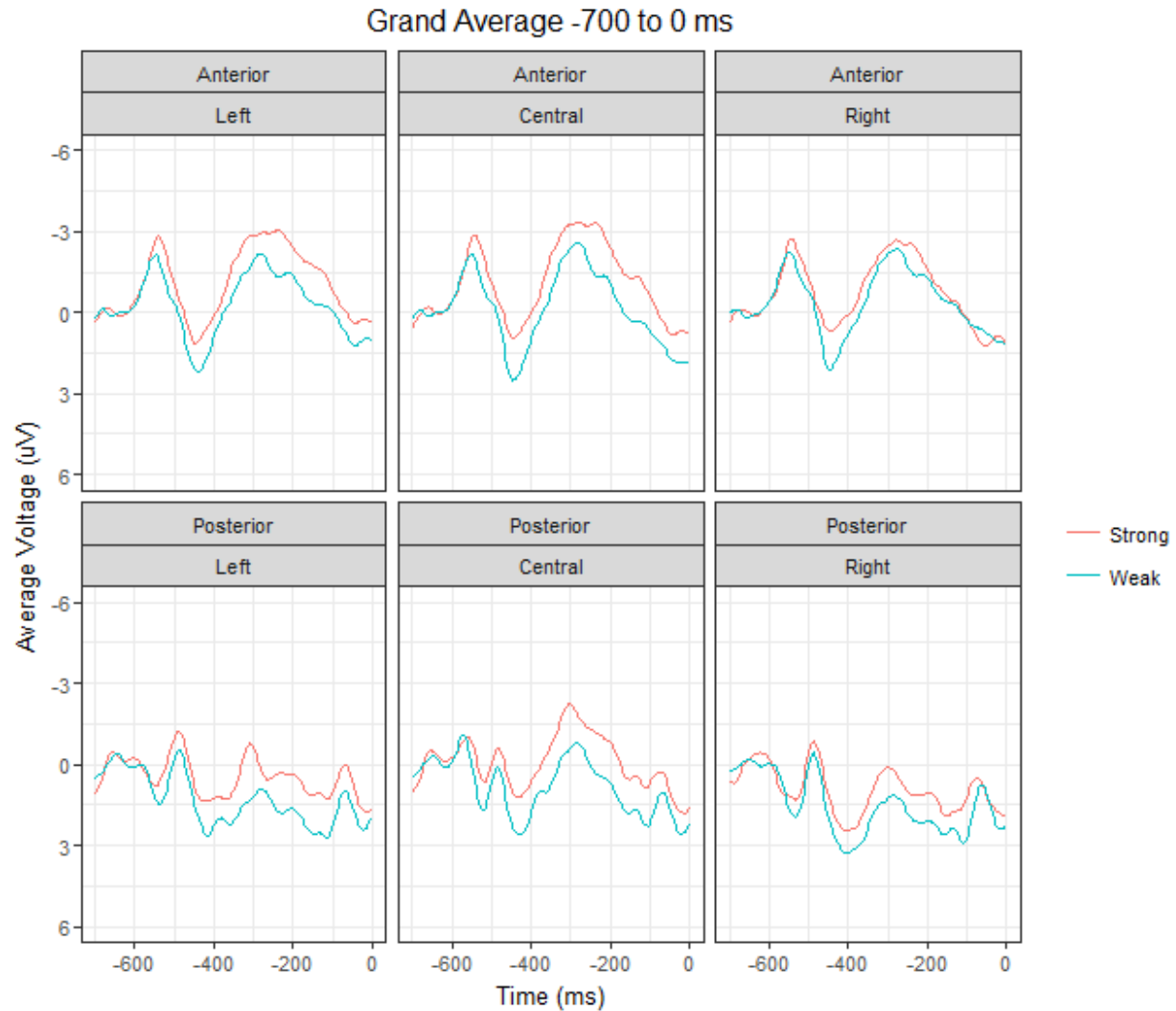
Figure 19 – Grand average of Escher data -700ms to 0ms.

Figure 19 displays a -700 to 0ms stimulus window with a -700 to -600ms baseline to avoid any possible effects that could be occurring early on. As expected, there are no statistically significant interactions in the C1 complex region after the onset of *Americans* or *American* for an equivalent window of -525 to -485 ms. This indicates that any effect this change in morpheme has is not an early recognition as indicated by previous data.

There are a few issues left in this study that will be addressed in the next experiment. Firstly, the data in this experiment should be interpreted cautiously, as there is

no control condition. The lack of a control condition means that it is unclear whether to call the observed deflections more positive or more negative. However, in spite of the absence of a control condition, post hoc analyses of Strong and Weak Eschers nevertheless provide evidence that these conditions are treated differently by participants.

Another limitation of the experiment was that there were no 'bad' fillers. The reason one would include a bad filler in a study like this is to provide a baseline rating comparison for what a low-rating would look like on a Likert Scale. However, even without the presence of bad fillers, participants still rated Weak Eschers markedly worse than Strong Eschers. A repeat of this experiment with the inclusion of bad fillers would likely replicate these effects, though it is important to establish clear boundaries for participants when providing them with a Likert Scale judgment task.

Strictly speaking, a case could be made that without the presence of bad fillers and of a control condition, the observed difference between Strong and Weak Eschers is difficult to interpret with confidence. So, the next order of business is to determine whether or not the differences remain when these two limitations are removed, as they will be in Experiment 4. In experiment 4, as I will show, both the behavioral and electrophysiological differences between Strong and Weak Eschers are replicated when bad fillers and a control comparative are included in the research design.

3.4   Experiment 4 – EEG Clarification

Abstract

Experiment 4 is a partial replication of Experiment 3, where a grammatical control comparative and bad fillers are added. Additionally, finer grained EEG techniques were used, including the usage of electro-oculography (EGG) and correction techniques for blinks. 23 participants underwent the same RSVP EEG experiment as in Experiment 3. Behavioral results are partially replicated here, though with the distinction of no significant difference between the Strong Escher and Control conditions. This is suggested to be due to timing effects, where participants are more likely to accept a sentence in a timed setting. EEG results replicate an early deflection at the 75-115ms window for the Weak Escher condition, supporting the claim that the parser recognizes the unlicensed ellipsis of the Escher Sentence.

3.4.1 Methods

3.4.1.1 Overview

The goal of this experiment was two-fold: one, to replicate the previous EEG findings from Experiment 3, and two, to reexamine the behavioral and EEG responses of Escher Sentences with a baseline control comparison. While Strong and Weak Eschers behave differently both behaviorally and neurophysiologically, introducing a control may either alter these differences or provide a clearer picture of what kinds of effects are present. With respect to EEG, a baseline comparative control will help determine whether the deflection in Experiment 3 was positive or negative. A control condition will also help to see if acceptability judgments from Experiment 2 are to be replicated since Experiment 4 can now make a direct comparison between Strong, Weak, and Control conditions.

3.4.1.2 Stimuli

From Experiment 3, changes to the stimuli were minor, including making some more readable, fixing minor errors, etc. The larger changes include creating triplets from the doublets, where a grammatical control was introduced into the paradigm as demonstrated in (39).

(39)   a. More knights fought dragons than the peasants did because of the massive fires engulfing the countryside. (Strong)

b. More knights fought dragons than the peasant did because of the massive fires engulfing the countryside. (Weak)

c. More knights fought dragons than peasants did because of the massive fires engulfing the countryside. (Control)

In (39c) the control case is introduced, and it only varies differently from (39a) by using a bare plural DP instead of an overt DP in the Strong Escher condition. As discussed previously, the removal of *the* from the comparative DP completely rescues the sentence, creating the control comparative, though the determiner is likely not the main driving force for the reason Escher Sentences can fool the parser. As in Experiment 3, 30 of each experimental condition were used, totaling 90 experimental sentences per participant. A 1:1 ration of fillers to experimental sentences was also maintained, resulting in a total of 180 sentences per participant. In addition to maintaining all the previous standards of stimuli construction from Experiment 3, each sentence was constructed using the preterit to avoid any computational difficulties with a more complex tense like past perfect; this also levels the paradigm such that the target area verb is always *did.*

3.4.1.3 Participants

There were 23 participants, 11F, 12M, ages 19-21 with a mean age of 20.87. Students were recruited from MSU campus. I was not the instructor for any students to eliminate any instructor-experimenter bias.

3.4.1.4 Procedure

The procedure for Experiment 4 follows a similar protocol to Experiment 3. Participants were welcomed to the lab and asked to read over a consent form. Following this, they were apprised of the entire EEG process along with some educational aspects such as explaining what EEG was, how it works, etc. After answering any questions, they were sat comfortably in a chair in front of a computer while they were fitted for a cap and gel was inserted into each electrode. Once this was done, they were shown their own EEG waveforms as continued education as well as to comfort participants, since many

participants ask if there will be any kind of shock. They were instructed to then pay attention to the screen in front of them as words flashed in front of them. The Cedrus Button Box was not used for this experiment; instead, a numerical keyboard was adopted for ease of use. PsychoPy (Peirce, 2018) was also used instead of SuperLab to deliver stimuli.

Words were presented using an RSVP paradigm, with a 600ms SOA and 200ms ISI. After each sentence, the computer displayed the question, "how acceptable was that sentence?" with a 1-7 reference on the screen, where 1 indicated "not a natural sentence of English" and 7 indicated "a completely natural sentence of English."

Data were recorded using a 32 sintered Ag/AgCl electrode cap (GND WaveGuard 64 Electrode cap; Advanced Neuro Technology BV., Enschede, The Netherlands). The data were amplified using a Full-band EEG DC Amplifier (Advanced Neuro Technology) with a sampling rate of 256 Hz. Both right and left mastoids were used as a reference, and all signals were captured continuously.

The same ROIs from Experiment 3 were used here, defined as— Anterior Left (Fp1, F7, F3, FC5), Anterior Right (Fp2, F4, F8, FC6), Anterior Central (Fz, Fpz, FC1, FC2), Posterior Left (CP5, P7, P3, O1), Posterior Right (CP6, P4, P8, O2), and Posterior Central (CP1, CP2, Pz, POz, Oz). The usage of electro-oculography (EOG) with two bipolar sets, a horizontal EOG (HEOG) and vertical (VEOG), was also used in addition to the 32 sintered Ag/AgCl electrode cap. These EOG recorded following the same procedure as the other

electrodes and were later subtracted from the data using a Gratton-Coles correction

method (Gratton, Coles, & Donchin, 1983).

Data across participants were collected and analyzed using EEGLAB, a plugin from

MATLAB (Delorme & Makeig, 2004). The average ERPs were calculated across participants

for all three conditions, Strong, Weak, and Control. The measured epoch was marked 200

ms before presentation of the target verb and extended 1000ms post-presentation for a

total window of 1200 ms.

Contaminated epochs were automatically rejected and double-checked by visual

inspection. Automatic artifact rejection was set at ($+/-$ 150 μV), and with the Gratton-Coles

correction, the rejection rate for bad trials was approximately 1% trial rejection. Subjects

that were two standard deviations or more away from the mean rejection rate were

thrown out, though in this case, all participants were kept. The ERP waveforms were

quantified by mean amplitude measures in relation to a 200 ms pre-stimulus baseline. ROIs

and conditions were subjected to repeated measures ANOVAs using R (R Core Team,

2013).

3.4.2   Results

3.4.2.1 Behavioral

For the behavioral results, the same methodology was applied as in Experiments 2 and 3. To reiterate, each participant's data underwent a z-score transformation. This data was then averaged across participants and conditions. Figure 20 shows the mean responses for each condition across participants.
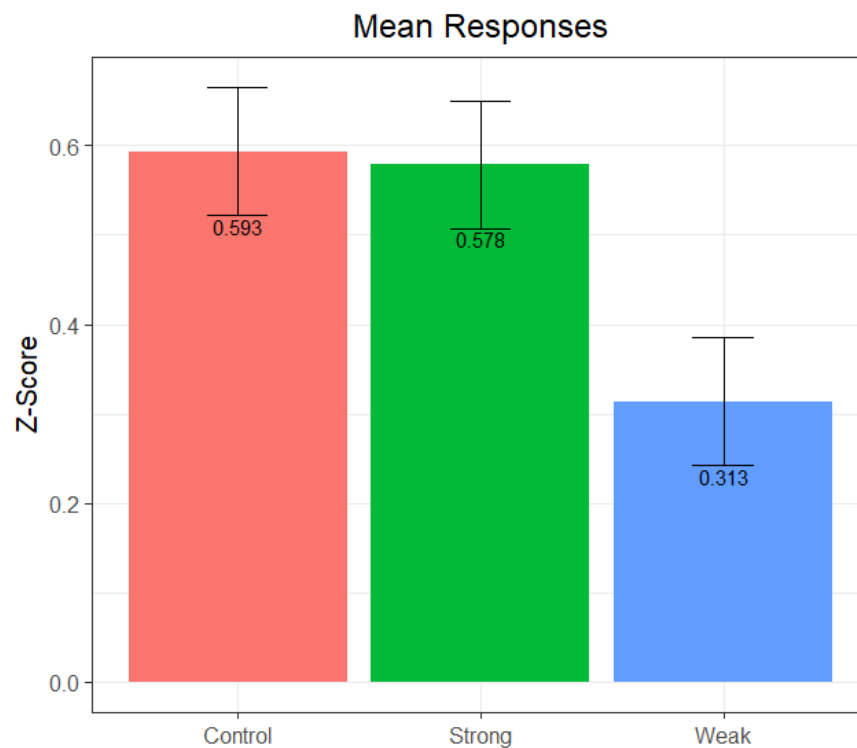
Figure 20 – Mean Responses for Control, Strong, and Weak Conditions.

Interestingly, in comparison with past experimental results, there was no difference between Strong and Control conditions, with a student's paired t-test indicating

[t(19)=.474, p=.641][12], though the difference between Strong and Weak remains where [t(19)=.5.395,p<.001*]. The distribution of responses for each condition is in Figure 21.
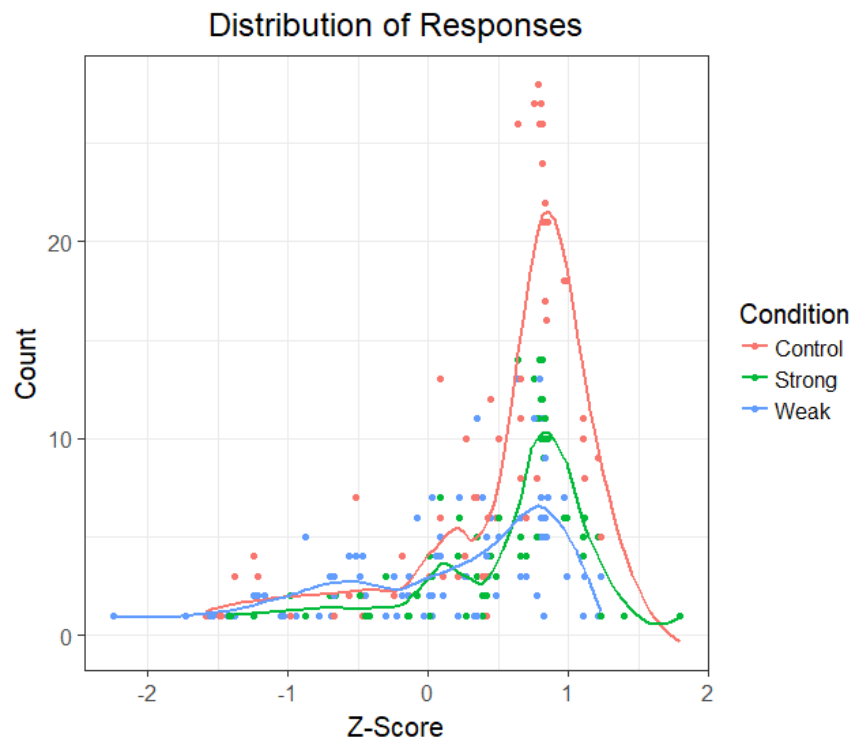
Figure 21 – Distribution of Control, Strong, and Weak Escher conditions.

In Figure 21, a one-way ANOVA revealed no significant difference between groups [$F(2,38) = 3.181$]. However, this ANOVA violated Mauchly's test for sphericity (Mauchly, 1940) (p=.048), so a Greenhouse Geisser Correction was applied (Greenhouse & Geisser, 1959), resulting in a p-value of p[GGc] = .067, just short of the conventional mark for significance of p <.05. The distribution of the fillers is also shown in Figure 22 to make sure

---

[12]Three participant's behavioral data are missing due to software failure; these missing data should not greatly impact the behavioral data trends here.

the fillers were doing their job of supplying a floor and ceiling for judgments for participants.
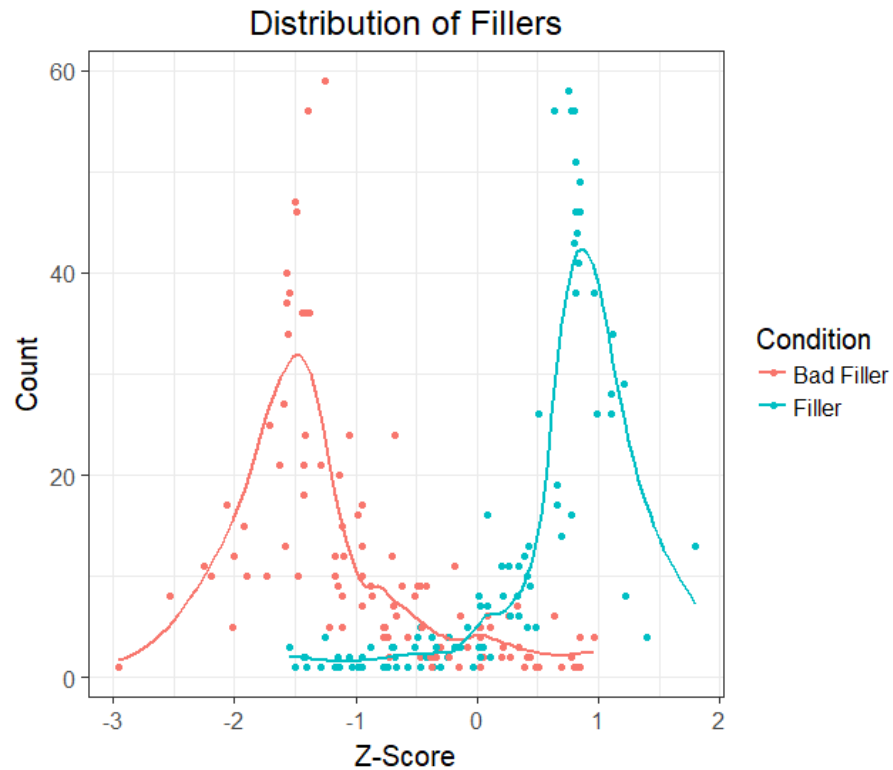
Figure 22 – Distribution of good vs. bad fillers for Experiment 5.

Unsurprisingly, a one-way ANOVA reveals a significant difference between groups [$F(1,19) = 68.438$, $p<.0001$], indicating that these two groups are quite distinct and act as clear markers for acceptable and unacceptable standards.

3.4.2.2 EEG

First, the mean amplitudes for each ROI and each condition were calculated and averaged across participants for a grand average, shown in Figure 23. Scalp plots of these data are also shown per condition in Figure 24, Figure 25, and Figure 26.
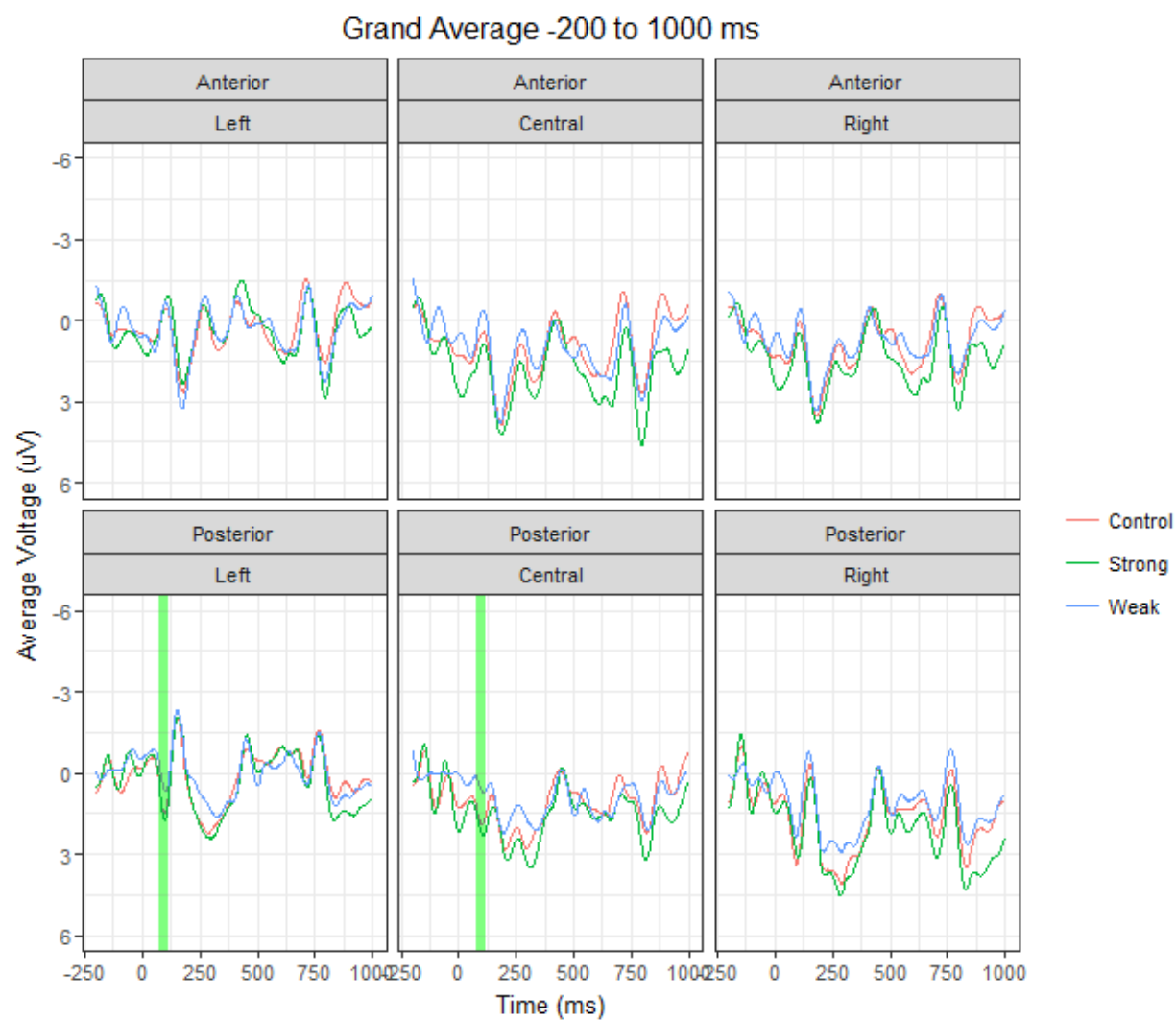
Figure 23 - Grand averages of Control, Strong, and Weak Escher in a -200-1000ms time window.
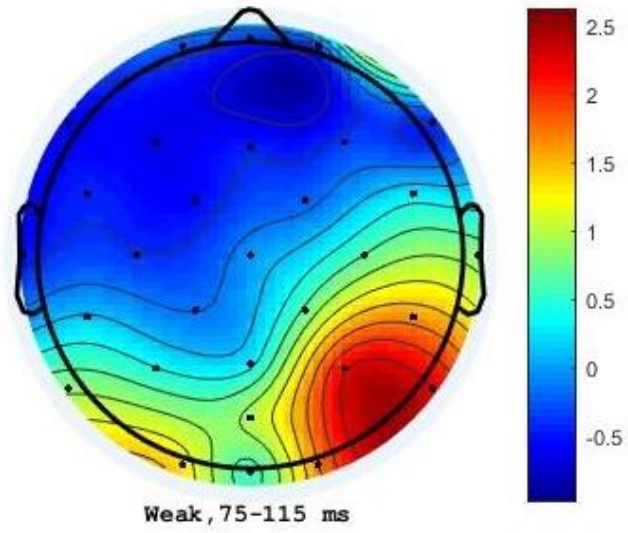
Figure 24 - Scalp plot of Weak Escher condition with mean amplitude between 75-115ms post stimulus.
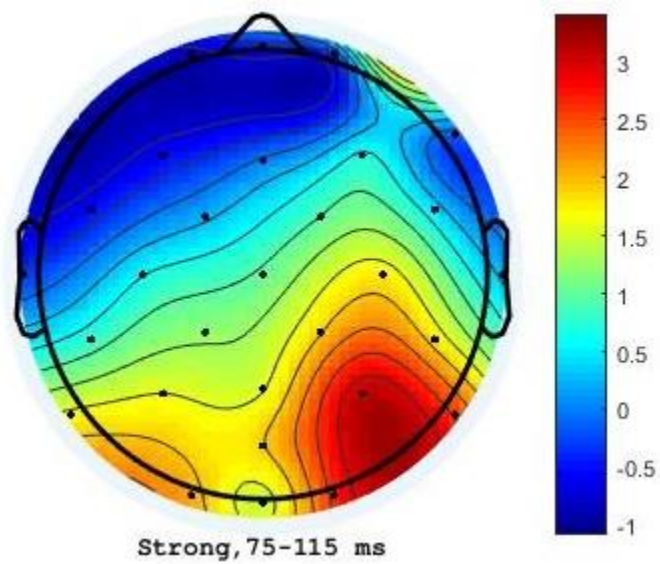
Figure 25 - Scalp plot of Strong Escher condition with mean amplitude between 75-115ms post stimulus.
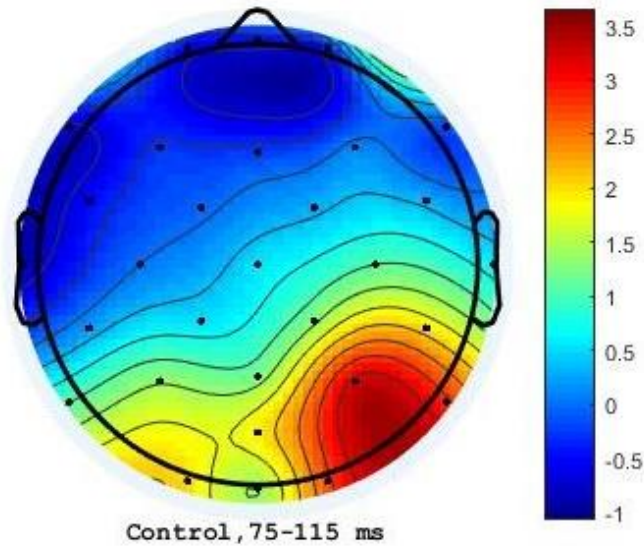
Figure 26 - Scalp plot of Control condition with mean amplitude between 75-115ms post stimulus.

In Figures 25-28, the window of interest here was 75-115ms, the same as in Experiment 3. Regarding the EEG data in Figure 25, there were main effects for Anteriority $[F(1,22) = 7.117, p=.014*]$ and Laterality $[F(2,44) = 7.414, p=.002*]$. There was significant three-way Anteriority x Laterality x Condition interaction $[F(4,88) = 2.732, p=.034]*$. No results violated Mauchly's test for sphericity. Because this interaction is significant, it licensed looking at any ROI with a post-hoc ANOVA. Post hoc analyses reveal significance in the Posterior Left Region $[F(2,44) = 3.503, p=.039]$ and approaching significance in the Posterior Central $[F(2,44) = 3.096, p=.055]$, though no significance in the Posterior Right Region $[F(2,44) = 1.260, p=.294]$. The regions of significance are again highlighted in the time course. Highlighted in scalp plots in Figures 26, 27, and 28, show a distinction in negativity in the posterior left region, where the Weak condition in Figure 25 has the lowest negativity in that region.

3.4.3   Discussion

The results from Experiment 4 replicate the early deflection 75-115ms after the

target area *did* for Escher Sentences. By introducing a grammatical control comparative

sentence to compare against Strong and Weak Eschers, the early deflection detected in

Experiment 3 becomes clearer. Here, in Experiment 4, this early deflection is likely a

negative-going deflection from the Weak Escher condition. This negativity is likely an

N100, which has been shown to modulate with changes to attention and discrimination

tasks (Luck et al., 1994; Mangun & Hillyard, 1991; Vogel & Luck, 2000). While these visual

tasks are not necessarily linguistic in nature, early effects for recognition of

morphologically complex words have also been shown in magnetoencephalography (MEG)

(Fruchter & Marantz, 2015; Solomyak & Marantz, 2009). Therefore, due to the family of

data surrounding the N100 as well as linguistic evidence of early syntactic work, it is likely

that the N100 can also visually track a linguistic error like that in the Escher Sentence.

With respect to the Strong Escher condition, however, behavioral data suggest

Strong Eschers lack any difference from controls. One possible explanation for this lack of

difference could be that participants were responding to these behavioral data in a timed

setting versus an untimed setting in Experiment 2. Perhaps the untimed setting allowed for

Escher Sentences to be less illusory in nature, and when in a more timed setting, perhaps

the illusory effect is stronger. I will seek to directly test what difference the timing of the

experiment can impact Escher Sentences in Experiment 5.

Another reason for a lack of behavioral difference between Strong and Control

conditions may be due to the frequency in which Number to Operation (NEO) is applied to

Strong versus Weak Eschers. In the Weak case, there is less illusory strength, making the

sentence more likely to be marked as unacceptable. In the Strong case, the illusion is stronger, which could mean that NEO is applied more frequently to these structures thus attenuating early effects and increasing their acceptability. NEO may be applied more frequently due to the comparative DP being closer to a bare plural DP than a singular overt DP as in the Weak case.

3.5   Experiment 5 – Event Biases and Necessary Ellipsis

Abstract

Experiment 5 was a behavioral study that examined the acceptability and interpretations of Eschers in timed and untimed settings. 40 participants over the course of two days were tested using a whole-sentence presentation paradigm (timed) and a survey (untimed) that probed their acceptability judgments of Eschers on a scale of 1-7 as well as a multiple-choice task for an interpretation. Results indicated that timing affects the acceptability judgments of Escher Sentences, rendering Strong Eschers and Controls indistinguishable in an untimed setting and significantly different in a timed setting, in contrast to previously established findings. Judgment data also indicate a bias towards event-interpretations in an event-individual ambiguity. Furthermore, a steep drop in acceptability is noted when an Escher Sentence ellipsis site is filled in.

3.5.1   Methods

3.5.1.1 Overview

To frame Experiment 5, let us consider again the famous exemplar repeated below in (40).

(40)   More people have been to Russia than I have [been to Russia].

For an illusion to be successful in fooling a participant into accepting an ungrammatical sentence like the Escher Sentence in (40), two features must be present: one, a comparative construction with an event-individual ambiguity, and two, an unlicensed VP ellipsis in the comparative clause. If one of these components is missing, the sentence either crashes or the illusion is less successful. Consider (41).

(41)   a. *There are more people in Russia than I have [been to Russia].

        b. *More people have been to Russia than I have been to Russia.

(41a) is missing the crucial matrix clause ambiguity, forcing the reader to think about the number of people, and so the continuation of the sentence, regardless of ellipsis, fails. In other words, by removing the matrix clause ambiguity, the ability for the Escher Sentence to fool a participant likely drops. Likewise, in (41b), to pronounce the elided material seems to make the Escher Sentence less likeable, and this has been shown in previous experimental findings (Fults & Phillips, 2004). However, Fults and Phillips compared a sentence like (41b) against a sentence like (40) with blank space as their

comparative target area. This could then mean that their findings came as a result of participants preferring shorter and less redundant sentences (Rooth, 1993). However, if an acceptability comparison were to be made by comparing a sentence like (41b) with a sentence that also contains roughly the same amount of written text, then the difference between elided and non-elided Escher Sentences may become clearer.

In interpreting sentences like (40), there is also conflicting data from previous experimental findings in this dissertation, namely that in some cases Strong Eschers are rated significantly worse than a control condition, but in others, like Experiment 4, Strong Eschers and Controls are rated almost the same way. One additional test that Experiment 5 establishes is to see if the differences in rating Strong Eschers and Controls is due to the timing of the task: if participants are forced to quickly respond, how will the strength of the illusion of an Escher Sentence be affected?

Hence, for the final two experiments of this dissertation, I aim to attribute the cause of the variable nature of Escher Sentences to their structure. To elaborate, an Escher Sentence is variable because of 1) a semantic ambiguity between events and individuals biased towards events and 2) an illicit ellipsis. To show this, Experiment 5 seeks to test if 1) there is a bias for interpreting ambiguities between events and individuals as events, and 2) if filing in the ellipsis site makes Escher Sentences degrade in acceptability. If there is a bias towards events, then it is likely the case that this is motivation for why NEO is inappropriately applied to Escher Sentences, and if ellipsis is important for Escher Sentences, then Escher Sentences should "stop working," or be much more clearly ungrammatical if the ellipsis is taken away. Moreover, I will introduce timing as another

variable to these data to show if time has some impact in how acceptable Escher Sentences are rated according to participants.

3.5.1.2 Stimuli

First and foremost, this experiment aimed to test several aspects of Escher sentences, so the stimuli had to reflect these needs. There were four conditions in this experiment, Control, Escher, Extended, and Non-Ellipsis, shown below in (42).

(42)   Control – More ogres ate trash than donkeys did.

Escher – More ogres ate trash than the donkeys did.

Extended – More ogres ate trash than the donkeys did which made the swamp a little cleaner.

Non-Ellipsis – More ogres ate trash than the donkeys ate trash which made the swamp a little cleaner.

The Control condition is a grammatical comparative sentence with *more* taking scope at the beginning of the sentence. The Escher condition here is a Strong Escher, where the predicate is repeatable and the comparative DP is plural with an overt DP. This pairing was designed to replicate previous findings and to determine if observing Escher Sentences in a timed vs. untimed setting affected how participants rated them. This condition will also help in determining, with respect to interpretations, if participants are more likely to think about ogres eating trash (events) or directly comparing the number of ogres and donkeys (individuals).

147

The second set of conditions, Extended and Non-Ellipsis, directly test if ellipsis is crucial for an Escher Sentence to be an Escher Sentence. The Extended condition adds ends of sentence correction after the elision site at *did*. The Extended condition will be compared with the Non-Ellipsis condition to see if the presence of ellipsis affects acceptability ratings. If the ratings of Non-Ellipsis are significantly lower than that of the Extended condition, this would be evidence to suggest that ellipsis is an important part of making an Escher Sentence illusory. Table 4 below summarizes the comparisons and predictions for this experiment.

Comparison	*Acceptability*		*Interpretation*	
	Timed	Untimed	Timed	Untimed
Control vs. Escher	Control>Escher	Control<Escher	Event	Event
Extended vs. Non-Ellipsis	Extended>Non-Ellipsis	Extended>Non-Ellipsis	Event	Event
Timed vs. Untimed	Timed > Untimed		Event	

Table 4 – A summary of the comparisons and predictions for Experiment 5.

Acceptability was determined as they have been in previous experiments, where participants encountered a sentence and rated it using a Likert Scale of 1 to 7. In the timed setting, judgments were gathered during whole sentence presentations using PsychoPy (Peirce, 2018) where the window to respond to each sentence was 10 seconds. In the untimed setting, judgments were gathered on paper surveys. For interpretations, each sentence was given a series of multiple choice responses where participants were asked *how do you interpret that sentence?* and their choices were recorded in both timed and untimed settings. Each set of multiple choice responses provided the reader with the

possibility of responding to the sentence as a comparison of events, individuals, or that the sentence had no interpretation. For example, if the sentence was *more ogres ate trash than donkeys did* (Control), their possible choices were — *(1) ogres ate trash more than donkeys did (Events), (2) There were more ogres than donkeys (Individuals), (3) No Interpretation, (4) Not sure.*

Sixty quadruplets of stimuli like those in (42) were constructed using a Latin Square Design. Each subject saw each condition at least once, resulting in a within-subjects design. Importantly, each participant saw 15 of each condition, resulting in 60 experimental sentences. Participants also saw 120 fillers, resulting in a 2:1 filler to experimental ratio. This increase in fillers was to further ensure that participants were not aware of any patterns in the data. For the fillers, half were acceptable fillers (Good Filler) and half were not (Bad Filler). For the Good Fillers, one quarter contained *more* at the beginning of the sentence, one quarter contained some comparative element, and the remaining half contained declarative sentences without comparisons. For the Bad Fillers, there was a mixture of comparative and declarative sentences, but all were ungrammatical to the extent that either words were missing or shuffled around the sentence.

These stimuli were delivered over the course of two days, where half the participants saw the timed version first and the untimed version second and vice versa to account for any affects this ordering might have had. Additionally, participants saw the same set of stimuli randomized differently between testing sessions. This was done to minimize the variability of their answers and to reduce possible learning effects of seeing the same stimuli in a 24-hour period.

3.5.1.3 Participants

    40 Participants, 25F, 15M, ages 18-66, with a mean age of 26.875, drawn from a

system at MSU called SONA (www.msu.edu/~annamc/SONA_-_Paid_Pool.html), were used

for this experiment. Age was one rejection criterion for these data, where two standard

deviations from the mean age were rejected, resulting in three participants being removed;

the mean age was 26.875 and the SD was 12.030. Ratings of Good and Bad Fillers were also

rejection criteria, where if a participant's average response to the acceptability of a good or

bad filler was more than two standard deviations away from the mean, they were removed.

For Bad Fillers, the mean rating was 2.711, with a SD of .7104, resulting in no rejections.

For Good Fillers, the mean response was 6.138 and the SD was .5811, and three

participants were below the two-standard deviation threshold, resulting in a total of 6

rejected participants.

3.5.1.4 Procedure

    Participants were first welcomed to the lab and instructed to sign a consent form

before beginning. They were instructed that they would be taking two tests, one today and

one tomorrow, and upon completion of the second task, they would be rewarded with $20

USD. This was done to ensure that participants showed up for both testing sessions. Half

the participants started Day 1 with the timed version and Day 2 with the untimed version.

    The timed version of the test was a whole-sentence presentation paradigm (see Lee

& Newman, 2009), where participants were presented with the entire sentence for 4

seconds. This was done to simulate reading in a timed setting without the necessary end of

sentence corrections required for word-by-word RSVP paradigms. Following this, each

sentence had two tasks, an acceptability judgment task as in previous experiments, where

the computer, after waiting 1 second following the sentence presentation, displayed *how acceptable was that sentence* with a reminder on screen of a scale of 1-7, where 1 was *not a natural sentence of English* and 7 was *a perfectly natural sentence of English*. This task timed out after 10 seconds if there was no response, and participants were encouraged to answer as quickly as possible. They were also instructed to use the keyboard buttons 1-7 to respond. Then, after another 1 second pause, the computer displayed *how do you interpret that sentence?* with a multiple choice below. They were instructed to respond using the keyboard buttons 1-4, where 1 and 2 were counterbalanced to be between an event interpretation and an individual interpretation. For example, if a participant encountered the following sentence—

> *More ogres ate trash than donkeys did.*

Their choices were as follows—

1) There were more ogres than donkeys (individuals).
2) Ogres ate more trash than donkeys (events)
3) Not sure.
4) No interpretation

Choices 3) and 4) were provided as options to indicate if they were not sure or if they truly thought the sentence was bad. Additionally, the acceptability task always came before the interpretation task because if a participant marked a *no interpretation* or *not sure* response, this could bias their acceptability of the sentence. Participants were given a

series of practice sentences before the start of the experiment to get used to the paradigm. For this task, all timing was recorded using PsychoPy (Peirce, 2018).

The untimed task had the exact same stimuli though in a different order. Here, the sentences were on a written survey, and participants were encouraged to take their time. Sentences were placed on a giant table, where acceptability scales and interpretation choices were clearly displayed at the top of each page. Participants were asked to fill in acceptability first followed by interpretation judgments, as acceptability boxes were to the immediate right of the sentence. Timing was recorded manually, when participants started to write their first response to when they completed the entire survey.

3.5.2 Results

3.5.2.1 Acceptability

As for all acceptability data gathered in this dissertation, each participant's data were gathered and z-score transformed. These results were averaged across participants and across conditions. Figures 27 and 28 show the mean responses for timed and untimed settings, respectively.
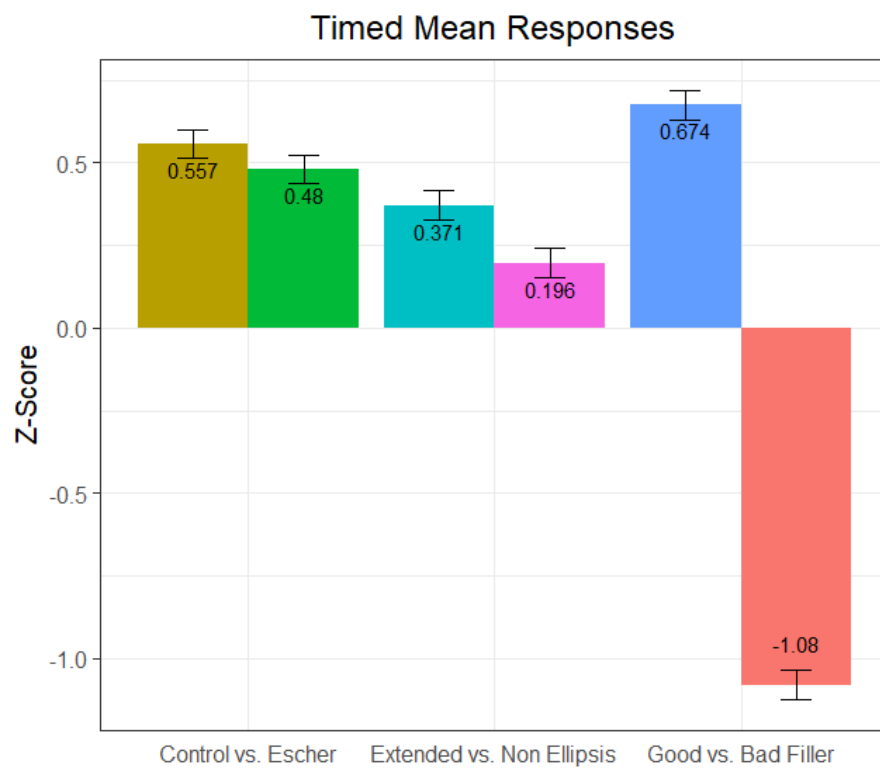
Figure 27 – Timed mean responses grouped by comparisons.
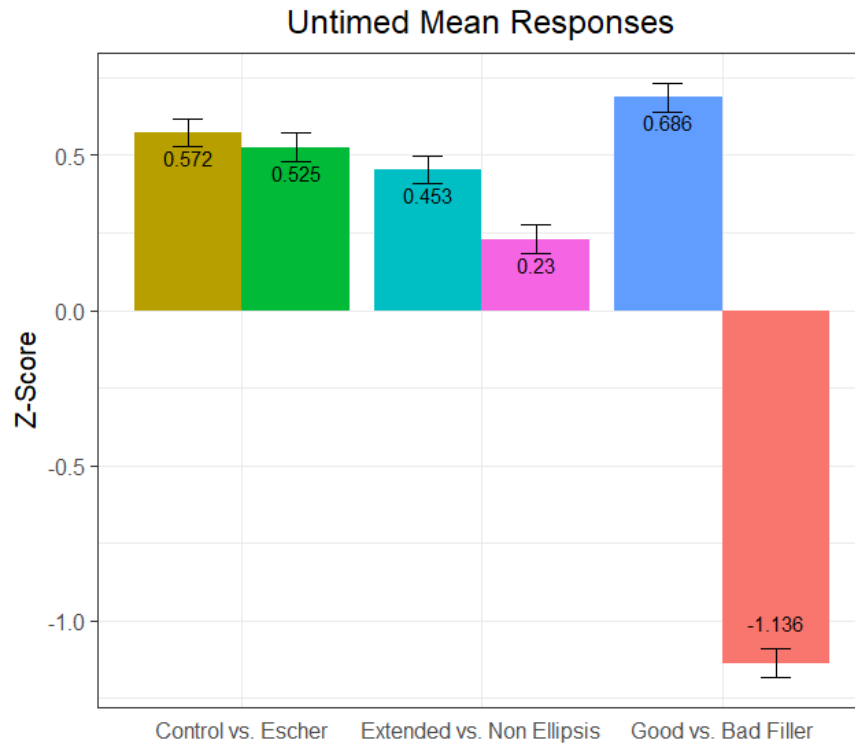
Figure 28 – Untimed mean responses grouped by condition.

For Figures 27 and 28, the most drastic differences are between Good and Bad Fillers; these conditions appeared to do their job for providing clear good and bad judgments for participants. Next, in both timed and untimed settings, there appears to be distinct differences between the Extended and Non-Ellipsis conditions. Table 5 shows paired t-tests of the overall time vs. untimed settings for each condition.

Comparison – Overall Time vs. Untimed	*Paired T Test*
Controls	$t(33) = -.528, p=.601$
Eschers	$t(33) = -1.577, p=.124$
Extended	$t(33) = -2.163, p=.038*$
Non-Ellipsis	$t(33) = -1.035, p=.308$
All Data	$t(135) = -2.779, p=.006*$
Experimental Sentences (No Controls)	$t(101) = -2.802, p=.006*$

Table 5 – Paired t-tests comparing timed and untimed responses.

First, comparing timed and untimed settings overall, untimed ratings are

significantly higher than in timed settings. All Data refers to every condition combined, and

No Controls removes the Control condition. However, Control, Escher, and Non-Ellipsis

show no difference between settings. Table 6 below shows comparisons between

conditions and between settings.

Comparison	Setting	Paired T Test
Control vs. Escher	Timed	t(33) = 2.304, p=.028*
Control vs. Escher	Untimed	t(33) = 1.643, p=.110
Escher vs. Extended	Timed	t(33) = 1.853, p=.073
Escher vs. Extended	Untimed	t(33) = 1.671, p=.104
Non-Ellipsis vs. Extended	Timed	t(33) = -3.750, p<.001*
Non-Ellipsis vs. Extended	Untimed	t(33) = -6.456, p<.001*
Escher vs. Non-Ellipsis	Timed	t(33) = 4.257, p<.001*
Escher vs. Non-Ellipsis	Untimed	t(33) = 5.424, p<.001*
Non-Ellipsis vs. Control	Timed	t(33) = -5.340, p<.001*
Non-Ellipsis vs. Control	Untimed	t(33) = -5.120, p<.001*

Table 6 – Paired t-tests comparing between conditions.

Here, of interest is that between timed and untimed settings, Eschers are not

significantly different from Controls in an untimed setting but are in a timed setting,

Between Non-Ellipsis and Extended conditions, in both timed and untimed settings, Non-

Ellipsis and Extended conditions are both significantly different.

Regarding the distributions of these sentences, Figures 29 and 30

show the distributions of answers in timed and untimed settings.
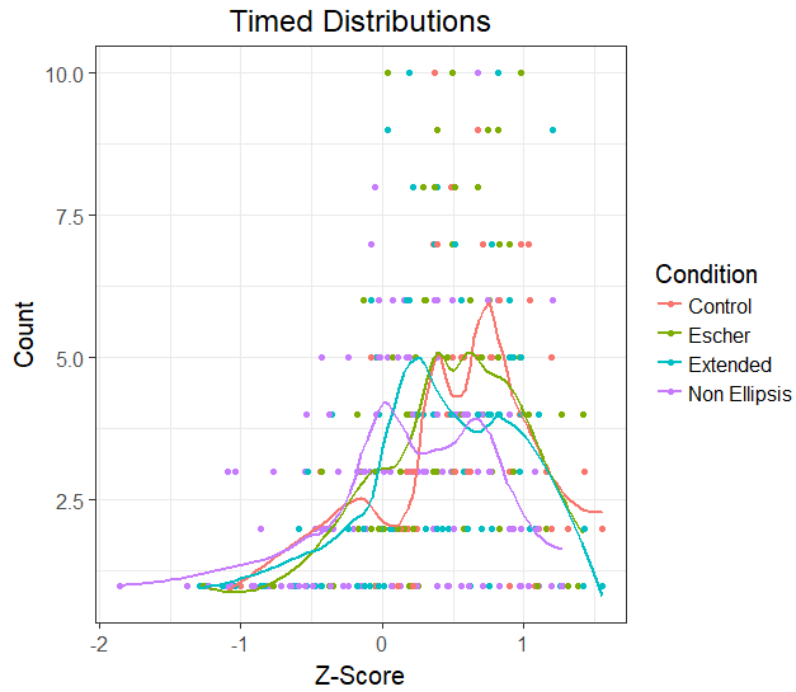
Figure 29 – Timed distributions of Control, Escher, Extended, and Non-Ellipsis conditions.
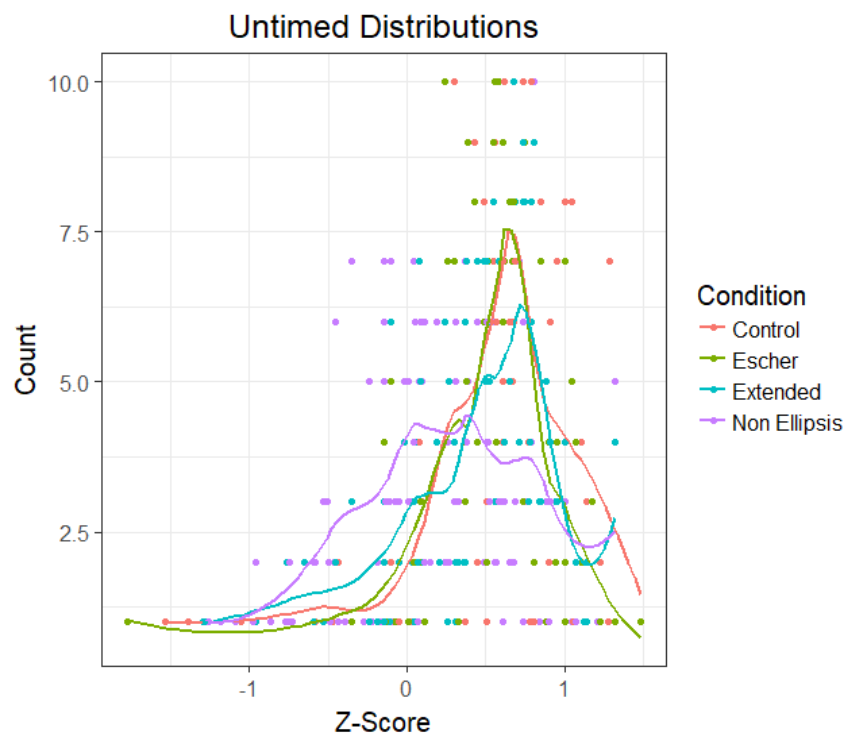
Figure 30 – Untimed distributions of Control, Escher, Extended, and Non-Ellipsis conditions.

One-way ANOVAs reveal significant differences in both timed and untimed settings between conditions, for timed [F(3,99) = 11.142, p<.001] and for untimed [F(3,99) = 10.691]. This untimed setting result violated Mauchly's test for sphericity (p=.028), resulting in a Greenhouse Geisser corrected p value of p[GGc] < .001*. The distribution of the fillers were also computed for both timed and untimed settings, which again show that the fillers provided a context for what to rate as good and what to rate as bad. These data are shown below in Figures 31 and 32.
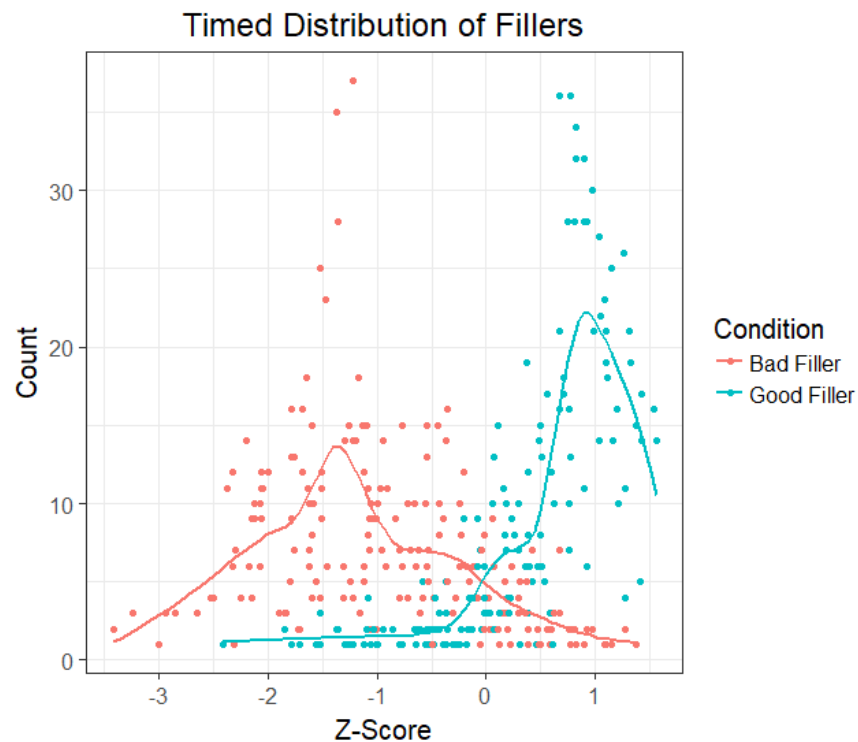
Figure 31 – Timed distribution of good and bad fillers for Experiment 5.
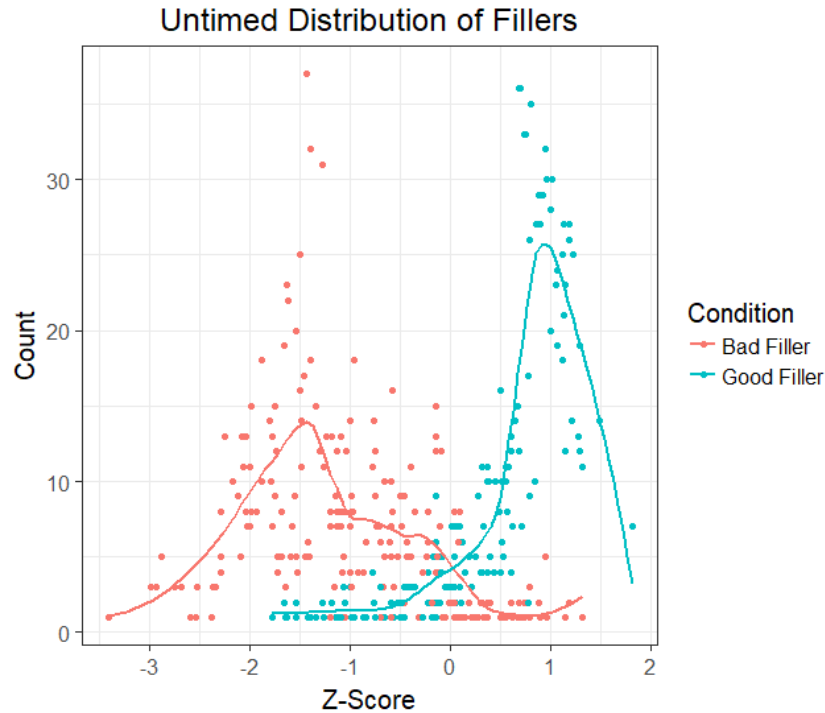
Figure 32 – Untimed distribution of good and bad fillers for Experiment 5.

Unsurprisingly, there are differences between fillers in both settings, for timed setting [$F(1,33) = 139.608$, $p<.001$] and [$F(1,33) = 220.096$, $p<.001$] in the untimed setting.

3.5.2.2 Interpretations

Regarding interpretations, data are reported as percentages of the total number of responses per condition. The proportion of selecting event for both Control and Escher conditions in a timed setting was 84.74%, while it was 97.41% in an untimed setting. To compare these percentages, the non-parametric Mann-Whitney-Wilcoxon Test (Mann & Whitney, 1947) was used. Since percentages do not follow a standard normal distribution, parametric statistical tests, such as ANOVA, cannot be used, and a non-parametric test must be used instead. The Mann-Whitney-Wilcoxon Test shows that there is a significant

difference between these timing settings [W=422, p=.019]. Individually, participants

selected an event interpretation for Controls 87.06% of the data in a timed setting and

98.82% in an untimed setting. For Eschers, participants selected event interpretations

86.47% of the time and 98.52% in an untimed setting.

For Extended versus Non-Ellipsis conditions, selecting an event-interpretation for

both conditions occurred 79.15% of the time in the timed setting and 95.91% of time in an

untimed setting. Per each condition, for the Extended condition, 83.53% of responses were

event-interpretations in the timed setting and 97.35% of data were event-interpretations

for the untimed setting. For the Non-Ellipsis condition, 83.82% of responses were event-

interpretations in the timed setting and in the untimed setting, 98.53% of the data were

event-individual responses. There was a significant difference between timed and untimed

settings between conditions [W=392.5, p=.011].

Overall, the probability to select event for any condition in a timed setting was

81.94% and 96.66% in an untimed setting, again reflecting a significant difference between

settings [W=1659, p=.001*]. Finally, overall, participants rejected 5% of the data, or they

selected *no interpretation* or *not sure* 5% of the time, meaning they found 95% of the data

interpretable.

Finally, Figure 33 shows the average time of completion for the timed and untimed

tasks. In the untimed tasks, the clock started when participants sat down and began

writing, and for the timed task, the clock started as the software began. The practice

session prior to the experiment adds to the timed task partially, but overall, the untimed

task took longer for participants to complete as they were instructed to take their time with

each question and were not forced to move on as they were in the PsychoPy stimulus

delivery system.
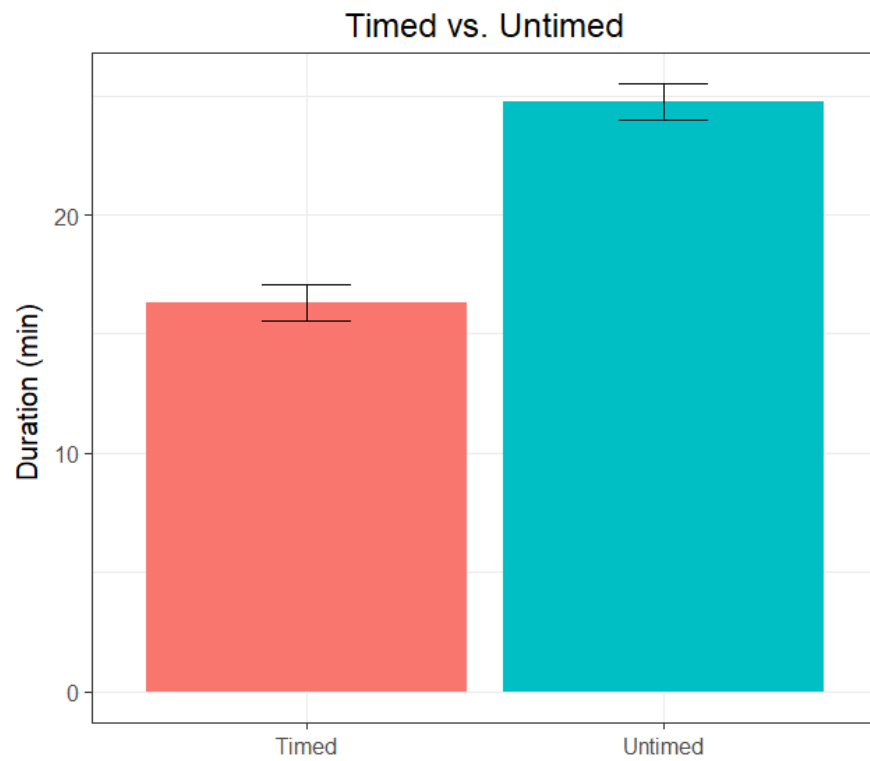
Figure 33 – Average duration of participant's time spent on Timed and Untimed tasks.

3.5.3   Discussion

These data provide a much deeper look into the behavior behind Escher Sentences. Specifically, they suggest that 1) regardless of being inside an Escher Sentence, a sentence with an ambiguity between events and individuals is likely to be interpreted as an event, 2) timing affects the interpretations and acceptability of Escher Sentences, where overall acceptability increases as well as the likelihood in selecting for event in an event-individual ambiguity, and 3) when an ellipsis site is filled in for an Escher Sentence, the acceptability drops significantly.

Regarding accuracy, experimentation has shown that accuracy is reduced when timed constraints are implemented (Mancini, Molinaro, Davidson, Avilés, & Carreiras, 2014). Therefore, the data gathered here in the timed setting versus the untimed setting could be due to accuracy issues. However, because these are judgment tasks, it is unclear how to gauge accuracy for sentences like Escher Sentences that have no actual interpretation and are so varied in their acceptability.

One problem to be addressed in the next experiment is the stark difference between the Extended vs. Non-Ellipsis conditions observed in this experiment. One idea to explain this difference is that people prefer non-redundant readings and may be marking the filled ellipsis sites worse than elided sites due to this effect (Rooth, 1993). However, as I will show in the next experiment, this difference is most likely negligible, as a control sentence with filled in ellipsis compared with a control sentence with the ellipsis intact shows little difference.

Another issue that arises is how the untimed paradigm was delivered. While I aimed to use a different medium, namely, writing, this introduced a confound into the project, as

this is a completely different way of responding to the stimuli than in the timed task. The solution here would be to present the untimed version also via a computer, and to not force times, but to merely record how long participants spent answering each question with the sentence remaining on the screen. Eye tracking studies here could help to tease apart what exactly may be going on, but for the time being, this difference offers a possible confound for these results. However, because of the consistency of the acceptability results in both cases as well as the following experiment, these differences may be negligible, but only further study can confirm this.

3.6   Experiment 6 – Replication

Abstract

Experiment 6 sought to refine and replicate the findings of Experiment 5. The Escher condition was removed from the paradigm and replaced with a grammatical control containing a filled in ellipsis site. 21 participants took the timed paradigm portion of Experiment 5. The results replicated previous results. They also revealed a weak difference between Control conditions with and without ellipsis, indicating that the redundancy of ellipsis had little effect on previous or current results. Interpretations were again biased towards event comparisons.

The final experiment for this project offers two major contributions to my findings. First, it offers a replication of my previous results. Secondly, I will run the exact same study, and though there will be no untimed portion, having the timed portion alone should indicate that the redundancy effect on ellipsis does not greatly affect the acceptability of the sentences.

3.6.1   Methods

3.6.1.1 Overview

This experiment serves as a partial replication of Experiment 5. Importantly, Experiment 6 introduces a second control sentence to account for any redundancy effect that may have occurred in Experiment 5 and which conceivably compromised the results of the comparison between Non-Ellipsis and Extended conditions.

3.6.1.2 Stimuli

The stimuli used were nearly the same as those used in Experiment 5, with two distinct differences. First, only timed stimuli are measured here, and the number of stimuli were reduced to 40 quadruplets from 60 to shorten the experiment. Given previous results, these timed stimuli alone should be sufficient to show that a Control sentence that contains a filled in ellipsis site will not affect the comparison between Extended and Non-Ellipsis conditions. Secondly, the Escher condition was removed and inserted in its place was a new control, Non-Ellipsis Control (NE Control), shown in the paradigm below in (43).

(43)   Control – More ogres ate trash than donkeys did.

NE Control– More ogres ate trash than the donkeys ate trash.

Extended – More ogres ate trash than the donkeys did which made the swamp a little cleaner.

Non-Ellipsis – More ogres ate trash than the donkeys ate trash which made the swamp a little cleaner.

Here, the NE Control serves as a comparison with Control. The idea is that if there is no significant difference between Control and NE Control, then the difference found in Experiment 5 between Extended and Non-Ellipsis can be regarded with greater confidence. Additionally, Non-Ellipsis and Extended are repeated here for replication purposes.

3.6.1.3 Participants

Experiment 6 participants were drawn from MSU classrooms. 21 Participants, 16F, 5M ages 18-22, with a mean age of 19.95, were recruited for this experiment. They were compensated with extra credit in their classes. I was not the instructor for any student. The same rejection criteria were applied to these subjects as in Experiment 5, but no participants met this criteria. In this experiment, participants were found to mistakenly press 8 or 9 for their Likert Scale judgments; these data points were thrown out.

3.6.1.4 Procedure

The methodology for the final experiment in this project replicates the timed version of the previous experiment, where participants sat in front of a computer and ran through a timed series of sentence judgments. To recapitulate, participants were presented with the entire sentence for 4 seconds. Following this, each sentence had two tasks, an acceptability judgment task just like previous experiments, where the computer after waiting 1 second following the sentence presentation, displayed *how acceptable was that sentence* with a reminder on screen of a scale of 1-7, where 1 was *not a natural sentence of English* and 7 was *a perfectly natural sentence of English*. This task timed out after 10 seconds if there was no response, and participants were encouraged to answer as quickly as possible. They were also instructed to use the keyboard buttons 1-7 to respond. Then,

after another 1 second pause, the computer displayed *how do you interpret that sentence?* with a multiple choice below. This task also timed out after 10 seconds.

Participants were given a series of practice sentences before the start of the experiment to get used to the paradigm. For this task, all timing was recorded using PsychoPy (Peirce, 2018).

3.6.2   Results

3.6.2.1 Acceptability

Again, following the same methodology as Experiment 5, mean responses were
calculated across participants after a z-score transformation, shown below in Figure 34.
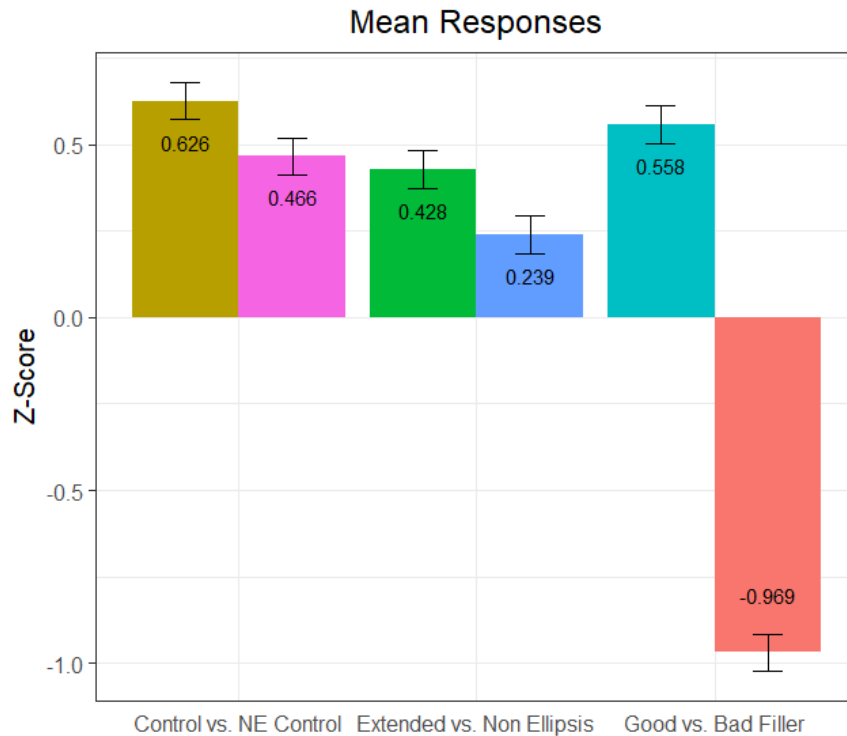
Figure 34 – Mean responses grouped by condition for Experiment 6.

As expected, the mean responses have the same shape and relationships as they
exhibited in Experiment 5, with the special exception of Control vs NE control, which did
not result in a significant difference, though it was close to significance. Importantly, the
difference observed in Experiment 5 between Extended and Non-Ellipsis conditions was
maintained. Table 7 below shows the relevant paired student's T tests.

Comparison	Paired Student's T-Tests
Control vs. NE Control	t(20) = 1.944, p = 0.066
Control vs. Non-Ellipsis	t(20) = -5.321, p = .00003*
NE Control vs. Extended	t(20) = 0.634, p= 0.533
NE Control vs. Non-Ellipsis	t(20) = 3.585, p = 0.002*
Non-Ellipsis vs. Extended	t(20) = -3.379, p = 0.003*

Table 7– Paired t-tests of comparisons between conditions.

Figure 35 below shows the timed distributions of the behavioral data.
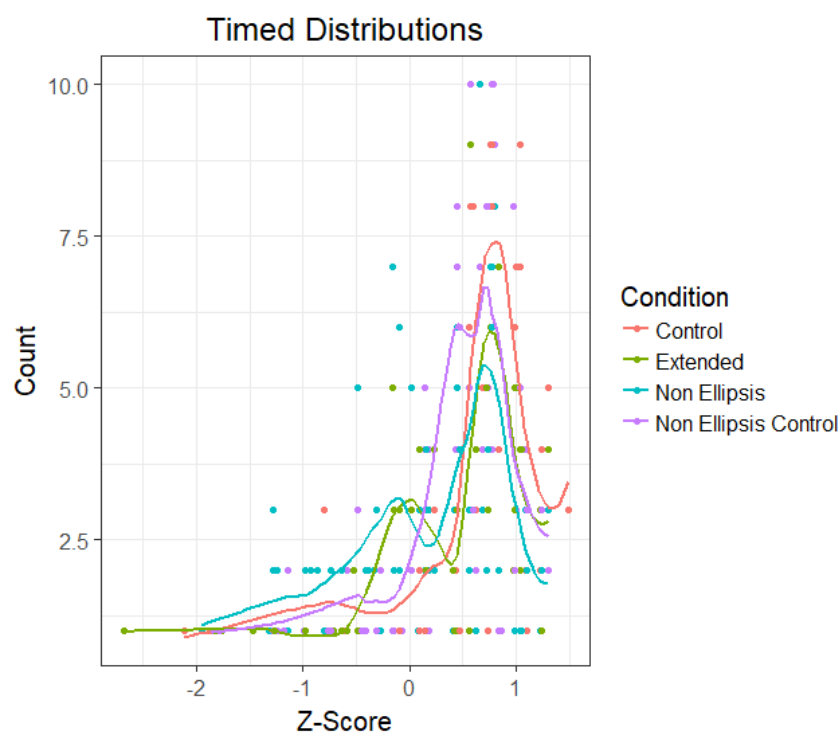
Figure 35 – Distributions of Control, Extended, Non-Ellipsis, and Non-Ellipsis Control conditions.

In Figure 35, a one-way ANOVA reveals no significant difference between conditions [F(3,60) = 2.321]. This ANOVA violated Mauchly's test for sphericity (p=.02), so a Greenhouse-Geisser correction was applied resulting in a p-value of p[GGc]=.107. Additionally, filler distributions were calculated, as shown in Figure 36.

168
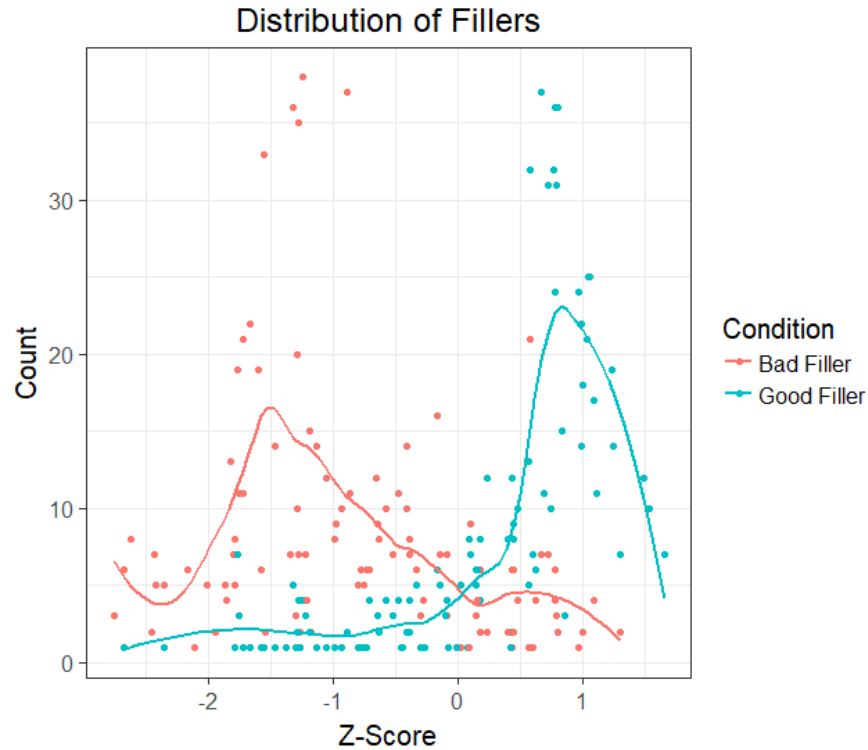
Figure 36 – Distribution of good and bad fillers for Experiment 6.

Again, as in all previous experiments, a one way ANOVA revealed a significant difference between conditions where $[F(1,20) = 26.089, p<.001*]$, indicating that fillers were useful in establishing high versus low acceptability ratings.

3.6.2.2 Interpretations

The proportion for selecting event for both Control and NE Control conditions was 53%. For Control 63.333% of responses were for event-interpretations and for NE Control 55.238% of responses were event-interpretations. For a comparison between Extended and Non-Ellipsis, the proportion of selecting events was 59.524% for Extended, 57.143% for Non-Ellipsis, and for participants selecting both Extended and Non-Ellipsis conditions as an event-interpretation was 52.666%. Overall, 16.5% of the data were rejected with *no interpretation* or *not sure* responses, leaving 83.5% of the data interpretable for

participants. While these data are not as stark as previous results, they are trending in the same direction as Experiment 5.

3.6.3 Discussion

Experiment 6 replicated some previous findings regarding the acceptability and interpretation behaviors for Eschers. Firstly, there was a replication between the Extended and Non-Ellipsis condition, with a lack of a statistical difference between a control case with ellipsis and one without; hence, any effect of redundancy can be disregarded in the previous experiment. Even if there was some effect for redundancy, it is likely the case that this played only a minor role in the difference between Extended and Non-Ellipsis conditions. I therefore take these results to indicate that without ellipsis, the effect of being fooled by these sentences is drastically reduced, consistent with claims made by AET.

Looking at the interpretation data in this experiment, while there is no untimed version, these data indicate that the majority favored response for participants is to interpret any sentence with a possible event-individual ambiguity as events rather than individuals. Furthermore, the lack of rejections in both experiments indicate that for the most part, participants are arriving at an interpretation when encountering Escher sentences, specifically an interpretation about comparing events. This gels with the AET and NEO, as participants are likely to use this strategy to arrive at an interpretation given the structure and semantic information available to the parser.

4.0   Discussion and Further Research

The final chapter of this project will revisit the experimental results and provide a summary of the main findings, followed by a discussion of the implications of these findings. I will discuss my findings with respect to the limitations of the parser, further support for a one-system view, a redefining of selective fallibility, and the refutation of the claim that the parser misses or ignores the error of an Escher Sentence. I will also elaborate on how these results could improve our models of artificial intelligence and natural language processing. Additionally, I will explore some unresolved issues, including the possibility of individual parsing biases, the nature of EEG and grammatical illusions, and the problems of random variation. I will then offer some recommendations for improving upon the work done here as well as calls for pushing the study of grammatical illusions and Escher Sentences further.

4.1   Discussion of Experimental Results

Experiment 1 was a large-scale behavioral survey of the interpretations of Eschers Sentences. The results of this survey indicated that there is a high degree of variability in the judgments of these sentences. The raw count of true/false responses revealed a pattern that participants were tracking event interpretations with Escher Sentences, and a lack of "not sure" responses indicates that participants are confident that Escher Sentences have some interpretation. However, upon further statistical examination using bivariate correlations, participants appeared to tracking individual interpretations with Escher Sentences. This suggests that while participants are confident in interpreting Escher Sentences, their interpretations are not consistent.

Experiment 2 probed at the acceptability judgments of Escher Sentences using a behavioral acceptability judgment task. Experiment 2 established a distinction between two classes of Escher Sentences, Strong Eschers, meaning more likely to be illusory or more likely to fool a participant into thinking they are acceptable sentences, and Weak Eschers, meaning less likely to be illusory. While other authors have noted that the changes in repeatability in the predicate and plurality in the comparative DP create noticeable differences in acceptability (O'Connor, 2015; Wellwood et al., 2009, 2017), this research replicated these effects, in addition to suggesting that Escher Sentences can be broken down into sub-classes

Experiment 3 examined Escher Sentences with an EEG paradigm. The behavioral data here replicated previous findings in that Strong Eschers are rated significantly higher than Weak Eschers. EEG evidence revealed an early deflection spread across the posterior

region from the 75-115 ms post-stimulus window. This deflection suggests early recognition of an unexpected token or error at the target verb *did.*

In Experiment 4, EEG data replicated the same early deflection 75-115ms across the posterior region, though it lacked significance in the Posterior Right region. With a comparative control as a baseline of comparison, the data suggest the effect was strongest in the Weak Escher condition. This deflection, when compared with the Strong Escher and Control conditions, was negative going. This suggests a likely N100 effect, where visual data are unexpected; I extend this to mean that when the parser encounters the target verb *did*, it recognizes that this ellipsis is illegal, thus eliciting possible Number to Event Operations (NEO) to correct or crashing. This variable response is shown in the distribution of the behavioral data for both Strong and Weak Eschers. Interestingly, there was no statistical difference between the acceptability judgments of Strong Eschers and Controls. This speaks to the nature of how Strong Eschers are much more likely to induce an incorrect NEO than in the weak condition, where participants were less likely to accept them. These EEG results thus indicate that the parser is not always blind to the errors of Escher Sentences,

Experiment 5 was another behavioral judgment task, though it probed at several aspects of the Escher Sentence. It contained a timed and untimed survey where participants looked at the same data 24 hours apart. Within either setting, participants were tasked with judging the interpretation and acceptability of Escher Sentences. Behavioral data in both timed and untimed settings revealed a large degradation in acceptability judgments between Escher Sentences with and without ellipsis, suggesting an importance of the presence of ellipsis. With respect to timing, data across the board were

rated more highly in the untimed setting than in the timed setting. Accuracy has been noted to be reduced in cases of time pressures (Mancini et al., 2014), which makes these data particularly interesting. If all judgment data is rated more highly, is this more inaccurate as some stimuli are ungrammatical and should be rated lower, or are untimed judgments more accurately reflecting the behavioral judgments of participants. Nevertheless, behavioral data with respect to interpretations in both timed and untimed settings indicate that the parser is inherently biased towards interpreting events in a context where there is an ambiguity between events and individuals. This was true for both Escher Sentences and control conditions.

Experiment 6 was a replication of the timed setting data from Experiment 5. The ellipsis data from Experiment 5 could have been potentially affected by the redundancy of filling in an ellipsis site. So, a grammatical control with a filled in ellipsis site was added to the stimuli set. The data regarding ellipsis are held from Experiment 5, as participants showed only a marginal decrease in acceptability in the control condition with a filled ellipsis site. Furthermore, timed acceptability and interpretation data were replicated in Experiment 6, bolstering the claims made by Experiment 5.

The data from Experiments 5 and 6 therefore suggest the resolution of event-individual ambiguities paired with illegal ellipsis may be the closest solution to explaining their cause and effects. The solution that many participants encounter with Escher Sentences is to think of them as event comparisons, reinforcing previous claims (Phillips et al., 2011; Wellwood et al., 2017). However, where Experiments 5 and 6 diverge from Escher Sentences simply being about events is that Escher Sentences also require ellipsis for the illusory effect to occur.

4.2   Implications of Results

The results gathered from this dissertation affect several aspects of cognitive and linguistic theory. This section will elaborate on the parser biases, support for a one vs. two system view of the parser, selective fallibility, and illusory strength. For convenience, the findings of this project can be summarized as follows—

- Escher sentences are ungrammatical but mostly acceptable sentences. These sentences are illusory, meaning that even though the structure is ungrammatical, many who encounter this sentence will accept it. EEG studies reveal that the parser recognizes a problem with the sentence, yet the resolution is not to always crash.

- Ambiguity and Ellipsis Theory (AET) is a hypothesis that explains the reason Escher Sentences do not behave like regular ungrammatical sentences. AET posits that Eschers are illusory because of an event-individual ambiguity interpretation in the matrix clause combined with an ellipsis that is unlicensed.

- Number to Event Operation (NEO) is a hypothesized operation that the parser may use to resolve an event-individual ambiguity.

- The parser is biased to interpret an ambiguity between events and individuals as events.

- The strength of the illusion in an Escher Sentence is greatly reduced if an ellipsis site is filled in or if the comparative DP is singular.

- The parser is biased to interpret an ambiguity between events and individuals as events.

4.2.1   Biases of the Parser

Escher Sentences highlight a distinct bias in the parser's ability to process ambiguity. Specifically, when encountering an ambiguity between an individual interpretation versus an event interpretation, such as a sentence like *three hundred dogs ran through the gate,* the experimental data from Experiments 5 and 6 suggest a bias towards resolving this sentence to be about events. This inherent bias contributes to the parser's mistake in incorrectly interpreting Escher Sentences as event comparisons rather than the null comparison available from the structure of the Escher Sentence.

In an ambiguity between reading something as an event versus an individual, in English, we are much more likely to resolve the ambiguity towards events. This could be for a variety of reasons. One avenue to explore would be to see how frequent event comparisons are made in English versus comparisons of individuals. A prediction here would be that constructing events from sentences is a more frequent structure in English. This would cue the parser into being biased in this direction, since while it is processing, if it predicts the more likely outcome, then there is less processing required.

If we assume that the parser tries to be as efficient as possible, then this hypothesis makes sense: select the interpretation that is the most frequent. The notion that our parser can be biased is already known, as there is evidence of other biases, e.g. in Garden Path Sentences. Here, the parser actively builds structures that are incorrect, "garden pathing" itself because of inherent biases towards simpler syntactic structures. Therefore, the findings regarding Escher Sentences continue a tradition that the parser is an inherently biased piece of machinery.

4.2.2 Support for a One vs. Two System View

To recapitulate, the divide in the psycholinguistic literature between a one-system and two-system view is based on how scientists model the grammar and the parser. In the one-system view, the grammar and the parser are one and the same, meaning that the mental mechanism that is responsible for computing grammatical information is the same device. On the other hand, the two-system view supports a view in which the parser is a device that processes language and accesses a separate entity, the grammar, and can rely on either the grammar for its computations or other cognitive heuristics. Escher Sentences provide a challenge for both theories. For the one-system view, the theorist must explain why ungrammatical sentences are in certain circumstances found to be acceptable considering that the parser has complete access to the grammar, being an aspect of the grammar itself, which should recognize that the Escher Semantic computation is null. For two-system theorists, they must explain what heuristics are used to both avoid the computation crash from the structure and to correct it.

The ERP evidence from Experiments 3 and 4 suggest that the parser recognizes an early error with Escher Sentences. This recognition may support a one-system view, though it could also be explained with a two-system view, i.e., the parser or some other mechanism has noted the Escher Sentence error. In the two-system story, the parser then variably accepts the sentence, less so with Weak Eschers and more so with Strong Eschers. The burden of explication on the two-system story, however, is to specify what operations are being used outside of the grammar. If we assume Number to Event Operation (NEO) is part of the grammar and is the typical operation for resolving event-individual ambiguities, then the variable responses that arise from the Escher Sentence are due to the nature of the

178

parser being biased towards event interpretations and the structure of the sentence varying in illusory strength rather than being due to a reliance on outside heuristics. Therefore, without assuming further heuristics, the Escher Sentence can be explained using a one-system view due to its recognition of an error and usage of an operation readily available in the grammar.

Thus, the questions shift away from heuristics and into the exact nature of the relationship between processing priorities and grammatical constraints. Clearly, some constraints are recognized and immediately dealt with, such as island violations, but in the case of illusions, specifically Escher Sentences, the problem is noted and variably dealt with. It may be best to think of this kind of variability not as variability between individuals, but rather as all individuals having the same parsing mechanisms and faced with the challenge of an incomputable sentence. If the parser encounters an Escher Sentence, there are only so many outputs it can create. In this case of Escher Sentences, the choices are either to crash or to move on, and we have seen that most of the time the parser appears to move on with a coerced event interpretation. Sometimes, participants do not accept Escher Sentences, and this variability of acceptance is like rolling loaded dice for the parser: it heavily leans towards events, but sometimes lands elsewhere.

This strategy makes sense from a Gricean communication standpoint. We assume a message makes sense until forced to conclude that it does not, and we do our best to assign an interpretation (Grice, 1975). If we heard something we cannot interpret, it would be ineffective to always abandon the attempt to arrive at an interpretation. We interpret as well as we can, so it makes sense that the parser should attempt to make sense of the non-sensical.

All in all, while the data from this dissertation does not warrant putting the argument of a one-system or two-system story to rest, Escher Sentences suggest no need to rely on the heuristics from a two-system story. Because the parser is able to recognize an Escher Sentence's error, and if the tools available to fix this error are part of the grammar, then there is no need to surmise additional steps to explain why participants are so variable with their responses to Escher Sentences.

4.2.3 Selective Fallibility and Illusory Strength

With respect to the concept of selective fallibility, or the notion that participants variably accept or reject grammatical illusions, the data from the Escher Sentence experiments reported in this dissertation have started the work of quantifying exactly how variable judgments of Escher Sentences are. While Escher Sentences represent a subset of grammatical illusions, it is likely the case that each grammatical illusion carries with it different levels of selective fallibility. To my knowledge, no experiment has yet directly compared several grammatical illusions together. However, in the definition of selective fallibility advanced by Phillips et al. (2011), is unclear whether selective fallibility refers to the parser or to the participant. In my research, I suggest moving away from placing the burden of variability on the participant and placing it on the parser instead. Thus, Escher Sentences, and likely grammatical illusions, are not variable because of participants choosing to accept them or not, but rather illusions are variable because of the nature of the parser and their structures. Therefore, a clearer concept to capture variability in grammatical illusions might be framed as Illusory Strength. Illusory Strength can be defined as a quantifiable concept that measures how likely a structure is to fool a parser into providing an inaccurate parse based on the structure of the sentence. In other words, if the sentence is ungrammatical, how likely is it to lead the parser into deriving an interpretation where there may not be one available.

Using the term Illusory Strength removes the burden of participants being selective or variable. Instead, it suggests that the parser and the structure of sentences are solely responsible for variability in their processed outcomes. Accordingly, it is highly desirable have a systematic way to categorize selective fallibility that could lead to predictions about

how a participant will behave with respect to any given illusion (Phillips et al., 2011). One way to approach the kind of predictive power required would be to reframe the problem. First, rather than describe illusions as sentences in which participants miss or overlook a constraint, let us recast the argument as the probability that the parser will incorrectly declare a structure as either grammatical or ungrammatical. The focus moves to the structure of each individual illusion rather than relying on individual variation. In this way, inquiry could begin to quantify the likelihood of a grammatical illusion succeeding or failing.

Exploring the limitations of the parser via the study of grammatical illusions helps shed light on an important aspect of the workings of the parser. All in all, humans are imperfect in their ability to comprehend and produce language. While many mysteries persist regarding the origins of language and how we acquire and use language so effortlessly, the performance systems are not flawless. Our parser is limited, as we have seen with Escher Sentences, but given these limitations, it is a matter of real interest to take a semantic/syntactic operation that yields no truth value and assign something to it. When we view the blue-black-dress, our eyes are not shuffling through every possible color combination in furious confusion, but rather the brain settles on something and moves on. Our language system operates in much the same way: it finds a troubling sentence and makes a parsing decision to accept it, something it is biased to do, and sometimes it crashes instead.

4.3   Recommendations and Further Issues

In this dissertation project, we have explored several aspects of Escher Sentences that have been used to support several claims regarding the nature of the parser, grammatical illusions, and linguistic theory. This section will lay out a few suggestions regarding where further research can take the topic of exploring grammatical illusions as well as issues that remain unanswered or require further investigation.

First and foremost, the ERP results from Experiments 3 and 4 suggest that the parser is able to recognize the problem with the illicit ellipsis in the Escher Sentence. The most likely ERP that encapsulates this recognition would be a visual N100. However, the work done on the N100 does not directly involve the recognition of early syntactic or semantic errors (Luck et al., 1994; Mangun & Hillyard, 1991; Vogel & Luck, 2000). However, some MEG evidence suggests early visual recognition of syntactic changes, so it is likely that the N100 can be modulated with discrimination tasks and attention as well as linguistic syntactic work (Fruchter & Marantz, 2015; Solomyak & Marantz, 2009). It is likely the case that participants are paying attention to something odd, namely an illicit ellipsis, for the Escher Sentence.

ERPs are still being mapped out and what each ERP indexes is not yet clear, so in order to support the claims made in this dissertation, there needs to be more study regarding the nature of what the visual system can identify at early processing stages. In the case of Escher Sentences, there is a clear deflection early in the EEG waveform which can be interpreted as a problem in the language itself. I suggest that this deflection marks a shift in attention to an unexpected stimulus which is the ellipsis of the Escher Sentence. If

further research can also support unexpected stimuli with respect to language, then this would have implications in understanding how the visual system and language interact.

With respect to grammatical illusions in general, another avenue to explore would be to see if one could divide participants into ideally two groups, those who likely accept grammatical illusions and those who do not. If we assume that these groups could be created, there could be valuable data to ascertain. For one, examining the neurophysiological reactions for someone who often cannot compute grammatical illusions would likely have large processing costs in an attempt to reconcile them, whereas the group of participants who readily accept grammatical illusions would likely have much lower costs in processing. This could also potentially lead into relating other cognitive functions that could affect processing grammatical illusions. So, for example, if the group of participants is likely to be fooled by grammatical illusions, how would this same group fare with classic visual illusions? Would this group be likely to see alternative colors to the blue-black dress, for example, more than the group of participants who reject grammatical illusions? One could also pursue other cognitive ability comparisons, such as visual-spatial skills, mathematical skills, etc. These comparisons could likely yield interesting findings with respect to cognition and our brains ability to handle illusions of all kinds in general.

Furthermore, studies that include other illusions compared to Escher Sentences is would be useful. In an effort to map out the Illusory Strength, or likelihood a grammatical illusion will succeed in fooling someone, one would need to run a series of experiments comparing the levels of acceptability between illusions. This could lead to a line of inquiry where illusions are ranked on their Illusory Strength, and this could give insight into what kinds of syntactic or semantic structures are difficult for parsers to compute. One concept

that remains unclear is what it is about the syntactic or semantic structures of other kinds of illusions that makes them behave the way they do, i.e., with variable responses rather than clear computations. Perhaps, if grammatical illusions could be ranked amongst each other, patterns could emerge that could give linguists a clearer idea of how grammatical illusions work in general and why some errors are so easily rejected, and others are not.

A challenge to comparing grammatical illusions with each other, however, would be that each illusion has different etiologies and structure. Thus, it would be difficult to, for example, identify a target area to observe processing costs across different illusions. Perhaps processing costs at a target area would be challenging, but inquiry could start with at least gathering comparative acceptability data across illusions as these would not require establishing a target area for a self-paced reading, eye tracking, or other behavioral study. Data gathered in this way would also help support a story of either variable participants or illusory strength.

As stated throughout this project, there has been a push to place the variability of judgments onto the parser and away from the individual. However, there are two alternative approaches to explaining the variability in responses to Escher Sentences. One would be to claim that each individual person has different parsing biases, and that some people's parsers are more likely to accept an Escher Sentence than others. Within a single language, there exists high levels of variation in cultural influences, slang, idioms, etc., which all occur in different frequencies. These inevitably could affect and train each human parser differently. For example, if person A grows up in an environment where double modal constructions are prevalent e.g. *I might could do that*, and person B grows up in an environment where these constructions are rare, then person A would be more likely to

interpret ambiguous strings as having multiple modals rather than person B's solution. Showing how these changes occur and mapping them out, however, would entail amassing a huge body of work.

The other direction one could approach this problem with would be to invoke the idea that the variability of Escher Sentence judgments stems purely from random variation. Generally, random variation in any experiment is difficult to account for, but the highly variable nature of these sentences calls into question of how to further address this issue going forward for this direction of study as well as for the field in general. Where exactly this random variation exists would be up for debate: it could exist in a experimental design, within each individual, within the grammar/parser, etc. Simply claiming that the problem of variability is solved with random variation yields a surface level answer, but this leaves a lot of room left to explain what exactly is random, why does this randomness exist, and is there any way to make a substantial prediction.

Finally, the data in Experiments 5 and 6 suggest that event interpretations in an ambiguous event/individual setting are preferred, though this claim should be tested further. First, it would be ideal to establish a corpus of frequencies of event comparisons and individual comparisons, and from there, a count of ambiguous event/individual comparisons ideally paired with some count of interpretation preference. Having these frequencies would strengthen the claim of an event bias. This idea of event bias also ties into the claim I made about NEO being more likely used for Strong Eschers than Weak Eschers, as Strong Eschers are more likely candidates for becoming an event interpretation due to their plural DP comparative than the Weak Escher with a singular DP comparative.

This is because if comparing individuals, it may be semantically odder to compare a group

with a singular person versus two groups, or two plural DPs.

4.4   Concluding Remarks

This dissertation serves as a one of few starting points for the investigation into Escher Sentences and grammatical illusions more generally. These illusory sentences are still largely a mystery for linguists and cognitive scientists, but by examining Escher Sentences in detail, we may be able to test the limitations and abilities of the human parser. I have provided evidence that suggests that Escher Sentences are not ignored by the parser. Instead, they are recognized and variably dealt with. This shows that the parser is more sensitive to grammatical illusions than previously thought. Moreover, evidence gained from Escher Sentences and event-individual ambiguous sentences shows that the parser is biased to resolve these ambiguities as events when it lacks context to inform it otherwise. I have also suggested that the cause of an Escher Sentence's illusory strength, or its ability to fool a participant, comes from an event-individual ambiguity combined with an unlicensed ellipsis. In the end, it is safe to conclude that humans are flawed communicators, but how our parser deals with these flaws is fascinating. Instead of acting like the computers developed in the past several decades, the human mind is taking the oddest data and making sense of it, though even these oddities can push our system to the very edge. Understanding this edge, however, allows us to examine ourselves with a more scientific lens, and in doing so, we can create something better than what we have.

APPENDIX

Escher Stimuli – Experiment 1

Scenario 1a

You are among a group of three knights, and you are discussing at dinner how many times you have seen the queen today. The knights have seen the queen once, where you have seen her six times, due to her need for sweets and wine that you must bring to her.

Scenario 1b

You are among a group of three knights, and you are discussing at dinner how many times you have seen the queen today. The knights have seen the queen six times, where you have seen her only once, due to the fact that she does not like squires very much.

Possible Statements:

1) More knights have seen the queen than just me.
2) The knights have seen the queen more than I have.
3) More knights have seen the queen than I have.

Scenario 2a

A professor and two other students are talking about a seminar that you have also taken. The professor comments that you have taken the seminar four times because of your interest in the subject matter, while the other students have only taken the seminar once.

Scenario 2b

A professor and several students are talking about a seminar that you have mutually taken. The professor comments that you have only taken the seminar once because of your lack in interest, while the other students have taken the seminar three times for easy credit.

Possible Statements:

1) More students have taken the seminar than just me.
2) The students have taken the seminar more than I have.
3) More students have taken the seminar than I have.

Scenario 3a

There are two ranchers talking about the cowboys down the road, and you are talking with them about how many times they have ridden a bull. You have ridden a bull seven times, while the cowboys are a bit afraid of the bull and have only ridden it twice.

Statement 3b

There are two ranchers talking about the cowboys down the road, and you are talking with them about how many times they have ridden a bull. You have only done it once, but the two cowboys have ridden the bull several times just this week.

Possible Statements:

1) More cowboys have ridden the bull than just me.
2) The cowboys have ridden the bull more than I have.
3) More cowboys have ridden the bull than I have.

Scenario 4a

You are talking with the owner of a mansion and two workers about who does more cleaning. Just today, you have cleaned the living room four times because the owner's son is such a slob, and the workers have only cleaned it once.

Scenario 4b

You are talking with the owner of a mansion and two workers about who does more cleaning. You have only cleaned the living room once today, but the workers have done it six times because of the mess the owner's son causes.

Possible Statements:

1) More workers have cleaned the room than just me.
2) The workers have cleaned the room more than I have.
3) More workers have ridden the bull than I have.

Scenario 5a

You and three nerdy friends are talking about who has seen the movie *Star Wars* more. You have seen it twelve times, as you are a big fan, and your friends have only seen it once.

Scenario 5b

You and three nerdy friends are talking about who has seen the movie *Star Wars* more. You have only seen it once, but your friends have seen it over a dozen times.
Possible Statements:

1) More nerds have seen *Star Wars* than just me.
2) The nerds have seen *Star Wars* more than I have.
3) More nerds have seen *Star Wars* than I have.

Scenario 6a

You are talking with your two friends about who has driven to Austin, Texas more times. You have driven there fourteen times, while your friends have only driven there twice.

Scenario 6b

You are talking with your two friends about who has driven to Austin, Texas more times. You have driven there only once, while your friends pass through there so often that they have lost count.

Possible Statements:

1) More truckers have driven to Austin than just me.
2) The truckers have driven to Austin more than I have.
3) More truckers have driven to Austin than I have.

Scenario 7a

You are working at a restaurant, and you are talking to two waiters about dealing with bad customers. Just today, you had to deal with seven bad customers, while they have only dealt with two today. It just is not your day.

Scenario 7b

You are working at a restaurant, and you are talking to two waiters about dealing with bad customers. You have only interacted with a single bad customer, while your co-workers have dealt with twelve bad customers. It is just not their day.

Possible Statements:

1) More waiters have dealt with bad customers than just me.
2) The waiters have dealt with bad customers more than I have.
3) More waiters have dealt with bad customers than I have.

Scenario 8a

You are attending a party with three actors, and you are talking about recent shows. You say that you have gone to see *The Book of Mormon* four times because you find it very funny, while your actor friends have only seen it once.

Scenario 8b

You are attending a party with three actors, and you are talking about recent shows. You say that you have seen *The Book of Mormon* once, but your actor friends have seen it twelve times because they know some of the people in the show.

Possible Statements:

1) More actors have seen *The Book of Mormon* than just me.
2) The actors have seen *The Book of Mormon* more than I have.
3) More actors have seen *The Book of Mormon* than I have.

Scenario 9a

Inside of the classroom, you and your two math friends are working on an equation. You have tried to solve the problem six times because are really struggling with this class, and even your friends have tried to solve it twice because of its complexity.

Scenario 9b

Inside of the classroom, you and your two math friends are working on an equation. You have tried to solve the problem twice because you are really struggling with this class, and, much to your chagrin, your math friends have tried to solve it six times because of its complexity.

Possible Statements:

1) More math students have tried to solve the equation than just me.
2) The math students have tried to solve the equation more than I have.
3) More math students have tried to solve the equation than I have.

Scenario 10a

A visit with your three friends leads to a heated argument about video games. You say that you have beaten *Super Mario* twelve times because of your skills, while your three gamer friends have only beaten it twice.

Scenario 10b

A visit with your three friends leads to a heated argument about video games. You say that you have beaten *Super Mario* only once because of your lack of interest, while your three gamer friends have beaten it twenty-six times, claiming it to be the best video games ever.

Possible Statements:

1) More gamers have beaten Super Mario than just me.
2) The gamers have beaten Super Mario more than I have.
3) More gamers have beaten Super Mario than I have.

Scenario 11a

You walk into a classroom where there are two ballerinas practicing. As you get to talking, you find out that you have seen Swan Lake over a dozen times because of your love of Tchaikovsky, while the ballerinas have only seen it once.

Scenario 11b

You walk into a classroom where there are two ballerinas practicing. As you get to talking, you find out that you have seen Swan Lake only once, compared to the ballerinas who have seen it six times because they wanted to improve their skills.

Possible Statements:

1) More ballerinas have seen Swan Lake than just me.
2) The ballerinas have seen Swan Lake more than I have.
3) More ballerinas have seen Swan Lake than I have.

Scenario 12a

While walking to class one day, you run across a couple of football players. You start to talk about workout routines, and you find out that you have run a marathon six times, while the football players have only ran a marathon once.
Scenario 12b

While walking to class one day, you run across a couple of football players. You start to talk about workout routines, and while you have run a marathon a few times, the football players have run marathons over a dozen times to keep in shape.

Possible Statements:

1) More football players have run a marathon than just me.
2) The football players have run a marathon more than I have.
3) More football players have run a marathon than I have.

Scenario 13a

You are at a local bar for trivia night with your friends, and they invite three girls over to play at the table. After some discussion, you say that you have won trivia night twelve times and are a reigning champion, while the girls have only won once.

Scenario 13b

You are at a local bar for trivia night with your friends, and they invite three girls over to play at the table. After some discussion, you say that you have won trivia night twelve times and are a reigning champion, while the girls have only won once.

Possible Statements:

1) More girls have won trivia night than just me.
2) The girls have won trivia night more than I have.
3) More girls have won trivia night than I have.

Scenario 14a

At the local library, you start talking with two of the librarians. You claim to have read *Harry Potter* seventeen times, while the librarians have only read it once. They then suggest other books for you.

Scenario 14b

At the local library, you start talking with two of the librarians. You tell them that you have read *Harry Potter* once, but they then comment that they have read it twelve times because they love the writing so much. You think that they should probably try reading different books once in a while.

Possible Statements:

1) More librarians have read *Harry Potter* than just me.
2) The librarians have read *Harry Potter* than I have.
3) More librarians have read *Harry Potter* than I have.

Scenario 15a

You are at a monastery and start to converse with three monks. The conversation turns to great books, and you say that you have read Augustine six times, while they have only read him once.

Scenario 15b

You are at a monastery and start to converse with three monks. The conversation turns to great books, and you say that you have read Augustine only once, while they have read his books on several occasions.

Possible Statements:

1) More monks have studied Augustine than just me.
2) The monks have studied Augustine more than I have.
3) More monks have studied Augustine than I have.

Scenario 16a

You are talking with two trainers at a gym. You say you have lifted weights six times this week, and they have only lifted weights once this week.

Scenario 16b

You are talking with two trainers at a gym. You say to have lifted weights only once this week in preparation. They themselves have lifted six times this week.

Possible Statements:

1) More trainers have lifted weights than just me.
2) The trainers have lifted weights more than I have.
3) More trainers have lifted weights than I have.

Scenario 17a

You are talking with two skiers about how you skied today. You fell down six times, while they admit that they both fell once.

Scenario 17b

You are talking with two skiers about how you skied today. You only fell down once, but the skiers both admit to falling down four times.

Possible Statements:

1) More skiers have fallen down than just me.
2) The skiers have fallen down more than I have.
3) More skiers have fallen down than I have.

Scenario 18a

You are talking with two carnival workers. You have ridden the rollercoaster twelve times today because you love the thrill, and the workers comment that they had to ride it once today to test it.

Scenario 18b

You are talking with two carnival workers. You have ridden the rollercoaster only once today, but you find out that the workers have ridden it six times to make sure it worked before the crowd arrived.

Possible Statements:

1) More carnival workers have ridden the rollercoaster than just me.
2) The carnival workers have ridden the rollercoaster more than I have.
3) More carnival workers have ridden the rollercoaster than I have.

Scenario 19a

You are working backstage at a theater and begin talking to two stagehands. You mention that you had missed an important cue seven times during rehearsal, while the two stagehands had only missed one cue.

Scenario 19b

You are working backstage at a theater and begin talking to two stagehands. You mention that you had missed one cue during rehearsal, while the two stagehands had missed several.

Possible Statements:

1) More stagehands have missed cues than just me.
2) The stagehands have missed cues more than I have.
3) More stagehands have missed cues than I have.

Scenario 20a

You are visiting the Napa Valley and attend a wine tasting event. There, you start talking with two wine connoisseurs. You mention that you have tasted the chardonnay three times, while the wine connoisseurs have only had it once today.

Scenario 20b

You are visiting the Napa Valley and attend a wine tasting event. There, you start talking with two wine connoisseurs. You mention that you have only tasted the chardonnay once today, while the wine connoisseurs have tasted it three times today, and they found it full bodied with an oaky note.

Possible Statements:

1) More wine connoisseurs have tasted the chardonnay than just me.
2) The wine connoisseurs have tasted the chardonnay more than I have.
3) More wine connoisseurs have tasted the chardonnay than I have.

Filler Questions

F1

You are in a group of friends talking about the latest video games. Some of your friends think that older games are better, but you feel that the newer ones have better graphics.

More people think older games are better.

F2

Someone spilled the milk in the refrigerator at work. A little annoyed, you decide to find the culprit. No one will fess up to the misdeed, though Dennis was the last person seen in the kitchen.

I think Dennis spilled the milk in the refrigerator a lot more than my coworkers.

F3

While on vacation in Florida, you find a group of beachgoers talking about a local restaurant. They say that the best crab legs are at The Crab Shack, but you think that the neighboring Lobster Lunch-in has the best shellfish in the county.

The Lobster Lunch-in has more variety than The Crab Shack.

F4

You notice that your roommate loves to leave dirty dishes in the sink. Frustrated, you confront him, but he ignores you. You start to leave sticky notes all over the dirty dishes, but you soon realize that this only exacerbates the tension.

There are more sticky notes on the dishes than I care to comment on.

F5

While writing your final paper for a class, you realize that you are missing almost 1,000 words. You also realize that the paper is due in an hour, so you are faced with a troubling situation.

You are missing more than 2,000 words for your final paper.

F6

Your friend and you are walking to the movie theatre, when you run into a stray husky by the side of the road. You both decide to skip the movie and take the dog home.

The dog was more of a husky than a Labrador.

F7

Listening to your favorite music, you decide that you need a little snack. You go into the kitchen and get some crackers and cheese dip. Unfortunately, on the way back to your room, you spill the cheese dip all over the new carpeting.

The rug ate more of the cheese dip than I did.

F8

You see a frog trying to cross the street one morning on your way to work. Even though you are in a rush, you decide to pick the frog up and carry it across the street. Feeling good about your deed, you then decide to give your employees a raise.

There was a pond across the street.

F9

After a long night of working, you decide to take a break in front of the television. Your partner, however, wants you to clean the living room. A bit annoyed that your TV time is cut short, you decide to do it anyway, though you only got to watch TV for 10 minutes.

There should be a lot more TV time after work is over.

F10

On your way to the gym one day, you see a couple of food carts that are selling freshly fried pickles. Having never tried one, you decide to visit one of the stands, considering you are about to work out anyway. Turns out that they are so good, you never make it to the gym.

More pickles were consumed than weights lifted on that day.

F11

Making dinner one evening, you realize that you have run out of garlic for your garlic mashed potatoes. As a substitute, you find a very old container of garlic salt, and though it said "Expires 2004," you desperately want those mashed potatoes.

Garlic salt offers a fine alternative to fresh garlic.

F12

At your favorite restaurant, you are torn between ordering the chicken or the steak. The waitress claims that the chicken is a better choice, but you look around and notice that a lot of people have steaks on their plates.

There are more steaks in the dining room than chicken dishes.

F13

At the Laundromat, you see a couple of friends doing some laundry. However, one of them forgot their quarters, but thankfully you brought some extra. Your friends were very grateful for the gift.

You didn't give your friends any quarters.

F14

You are in the mood to eat fresh-baked bread, so you head down to the local bakery. Disappointed at their hours, you leave empty handed and instead head to bed.

The bakery is open late at night.

F15

Watching the football game, you decide to go out and buy some more snacks. You ask your friends what they want, and they give you a variety of answers, but hot wings seems to be the winner.

Most of your friends like hot wings.

F16

You are out at the mall with some of your friends and you see a really cool jacket in the window. It costs well beyond what you can afford, yet you still really want it. Your friends offer a decent alternative, but it still is not the same.

Friends would prefer that you buy the expensive jacket more than the cheap one.

F17

You are visiting the zoo one day, and you notice that one of the rhinos has a top hat on. As you are setting up your camera to take a picture of the odd sight, the rhino bows down and tips his hat at you. Going to the zoo was a great idea.

The aquarium has the best exhibits.

F18

You are looking in to buying a new car. After finding the car that has the features you need at a good price, you are now thinking of the best color to match your personality. Green seems to fit you, but there is something special that you love about a red car.

You decided to buy a blue car at the dealership.

F19

Sitting at your work desk, you notice a strange bug near the corner that appears to be a walking stick. Excited, you run to find and show your coworkers, but by the time you get back, you see that it is a cockroach and Janet starts to scream. Deborah, on the other hand, tries to catch it in a bottle.

The cockroach on the corner of your desk scared Deborah more than Janet.

F20

Your friends are trying to decide which board game to play before heading out to dinner that evening. You recommend something simple like a card game, but Enrico wants to play a very complex board game. Though the reservations were at six thirty, you don't arrive at the restaurant until seven.

You played a board game with Enrico instead of a card game.

Escher Stimuli – Experiment 2

(1) More Brazilians made sandwiches than the American did, but everyone at the party enjoyed the food. (Weak)

More Brazilians made sandwiches than the Americans did, but everyone at the party enjoyed the food. (Strong)

More Brazilians made sandwiches than Americans did, but everyone at the party enjoyed the food. (Control)

(2) More knights fought invading dragons than the peasant did, and the village was luckily saved. (Weak)

More knights fought invading dragons than the peasants did, and the village was luckily saved. (Strong)

More knights fought invading dragons than peasants did, and the village was luckily saved. (Control)

(3) More Brits drank tea at the party than the American did, but good times were had by all. (Weak)

More Brits drank tea at the party than the Americans did, but good times were had by all. (Strong)

More Brits drank tea at the party than Americans did, but good times were had by all. (Control)

(4) More wizards casted spells than the sorcerer did, but the dragon continued to attack the town. (Weak)

More wizards casted spells than the sorcerers did, but the dragon continued to attack the town. (Strong)

More wizards casted spells than sorcerers did, but the dragon continued to attack the town. (Control)

(5) More magicians tricked people than the insurance agent did, but people were scammed regardless. (Weak)

More magicians tricked people than the insurance agents did, but people were scammed regardless. (Strong)

More magicians tricked people than insurance agents did, but people were scammed regardless. (Control)

(6) More nerds played board games than the jock did, but fun was had by all. (Weak)

More nerds played board games than the jocks did, but fun was had by all. (Strong)

More nerds played board games than jocks did, but fun was had by all. (Control)

(7) More cats caused mischief than the dog did, though nothing was badly broken. (Weak)

More cats caused mischief than the dogs did, though nothing was badly broken. (Strong)

More cats caused mischief than dogs did, though nothing was badly broken. (Control)

(8) More frogs jumped across the road than the raccoon did, and look how that turned out. (Weak)

More frogs jumped across the road than the raccoon did, and look how that turned out. (Strong)

More frogs jumped across the road than raccoons did, and look how that turned out. (Control)

(9) More grandmothers baked cookies than the mother did, but everyone enjoyed the baked goods anyway. (Weak)

More grandmothers baked cookies than the mothers did, but everyone enjoyed the baked goods anyway. (Strong)

More grandmothers baked cookies than mothers did, but everyone enjoyed the baked goods anyway. (Control)

(10) More zombies devoured villagers than the goblin did, and the king was mighty upset. (Weak)

More zombies devoured villagers than the goblins did, and the king was mighty upset. (Strong)

More zombies devoured villagers than goblins did, and the king was mighty upset. (Control)

(11) More cows grazed than the horse did, but the goats were left out of grass to eat. (Weak)

More cows grazed than the horses did, but the goats were left out of grass to eat. (Strong)

More cows grazed than horses did, but the goats were left out of grass to eat. (Control)

(12) More Americans swam in the race than the Brazilian did, even though the waters were cold. (Weak)

More Americans swam in the race than the Brazilians did, even though the waters were cold. (Strong)

More Americans swam in the race than Brazilian did, even though the waters were cold. (Control)

(13) More pigs rolled in mud than the dog did, but baths were needed by all. (Weak)

More pigs rolled in mud than the dogs did, but baths were needed by all. (Strong)

More pigs rolled in mud than dogs did, but baths were needed by all. (Control)

(14) More bikers ran the red light than the cab driver did, and the police were happy to give tickets. (Weak)

More bikers ran the red light than the cab drivers did, and the police were happy to give tickets. (Strong)

More bikers ran the red light than cab drivers did, and the police were happy to give tickets. (Control)

(15) More republicans ate pizza than the democrat did, but everyone enjoyed the meal. (Weak)

More republicans ate pizza than the democrats did, but everyone enjoyed the meal. (Strong)

More republicans ate pizza than democrats did, but everyone enjoyed the meal. (Control)

(16) More rafters packed cliff bars than the hiker did, as the trip required lots of energy. (Weak)

More rafters packed cliff bars than the hikers did, as the trip required lots of energy. (Strong)

More rafters packed cliff bars than hikers did, as the trip required lots of energy. (Control)

(17) More turtles swam in the river than the frog did, as the river has lots of food to offer. (Weak).

More turtles swam in the river than the frogs did, as the river has lots of food to offer. (Strong).

More turtles swam in the river than frogs did, as the river has lots of food to offer. (Control)

(18) More baboons played with the toys than the chimpanzee did, and the zoologists were intrigued. (Weak)

More baboons played with the toys than the chimpanzees did, and the zoologists were intrigued. (Strong)

More baboons played with the toys than chimpanzees did, and the zoologists were intrigued. (Control)

(19) More beauticians donated money than the barber did, but the charity appreciated all donations.

More beauticians donated money than the barbers did, but the charity appreciated all donations.

More beauticians donated money than barbers did, but the charity appreciated all donations.

(20) More birds ate seed than the squirrel did, though the feeder was empty at the end of the day. (Weak)

More birds ate seed than the squirrels did, though the feeder was empty at the end of the day. (Strong)

More birds ate seed than squirrels did, though the feeder was empty at the end of the day. (Control)

Fillers – Bad

1. The monkeys more mad at the zookeeper because of the food fight earlier in the day.
2. Cats higher than dogs jump all the time.
3. The pizza late got here and we were pissed off.
4. Even though it always done in the jump of time, things seem rather fishy.
5. Salads need something more badder than the other guy over there.
6. Jane the dog swim the cat over there by the fence in the backyard.

7. Try not to ate nails my grandmother would always say.
8. The amount of spillage the factory plant was not easy to clean up.
9. More than not do I need swim in river, I think it's a nice way of thinking.
10. Frank was surprised not me there to sway things in the right direction.
11. Archery practiced more by ancient folk than I care to think about.
12. Thinking about to what I need to think more than what needs thinking more.
13. What think about the bad times is not necessarily the yes.
14. Villagers saw taverns down to the ground burned, but the dragon was nary to be found.
15. Numbers can amount to monkeys more than keep at the zoo.
16. Dogs can never jump the river that river the jumped dog.
17. Fewer than twenty spells dragons hit the mark, but the villagers were envious of the wizards.
18. Digestion takes several oat cookies limited in the fact of internet told.
19. More than the average bear does the whale who not the swam the oceans blue.
20. The fewer I have to do about it, the best I cannot do about it.

Fillers – Good

1. More diners were able to compete with the larger chain restaurants, but eventually the larger corporations won out.
2. There were fewer amounts of poison found in the waters than anticipated, and the citizens rejoiced.
3. The zombies were annihilated by more peasants than knights at the end of the day, but regardless the village was saved.
4. Dogs tend to play a lot more than cats, but they both make perfectly good pets.
5. More than five hundred people signed the petition, which made the mayor reconsider the new law.
6. More lobsters were boiled at the cook-off than other fish of the sea were, but the judges love all kinds of seafood.
7. More musicians sold out as a result of money issues, but this is the nature of the business.
8. Fewer than 500 California Condors are left, which means we need to step up our conservation efforts.
9. More people named Robert have large foreheads than people named Steven do, but this could just be a made up statistic.
10. The scientists were more pleased at the progress of their experiments than the amount of money these tests are costing.
11. Fewer people think that global warming is fake nowadays, but many people still think it's a farce.
12. More kings are dethroned these days than the queens of other nations are, as the queens tend to rule more peacefully.
13. The less that I have to clean up after the party the better, because there is a lot of trash leftover.
14. Software develops spend way more free time coding than what their company often requires, but this extra work pays off in the long run.

15. More than often do I find myself browsing pages of Wikipedia, as the articles can be very interesting.

16. Less than enough people are doing something about the dragon attacking the town, and the king is very upset.

17. Golf requires more calories to play than bowling does, but they are both sports where enjoyment is more important than exercise.

18. Keyboards require more RAM than mice do, but most people use both to work on the computer.

19. The editor rejected more of my articles than I care to admit.

20. Losing weight is difficult for many Americans, but some people make it a lifestyle and succeed.

21. Many people think that pizza is a vegetable, but it is probably unhealthy to eat it every day.

22. The cats drink a ton of milk in the morning, but don't let the dogs nearby, they'll steal it all.

23. Butchers buy more meat than grocers do, but this is because of the nature of the job.

24. Many people grew up watching morning cartoons, but many didn't have that opportunity.

25. Digital clocks are outnumbering analog clocks more and more these days.

26. Headphones are a major cause of pedestrians walking into each other, but no more than people paying more attention to their cellphones.

27. The Amazon River is full of many varieties of wildlife, and many are dangerous to humans.

28. Movies tend to be longer in length if Peter Jackson directs them.

29. The shorter the time spent studying, the more time can be spent playing video games.

30. Fewer cats are able to withstand baths than most other pets, which is why they often just clean themselves instead.

31. Computers can be difficult for many people, but some people have a knack for it.

32. Rocket science is very difficult, and only a few people are able to study it.

33. Doctors prescribe more antibiotics than need be sometimes, and this practice is making bacteria even more resilient to drugs.

34. Green chilies are not as strong as jalapeño peppers are, but I tend not to like spicy food anyway.

35. Americans tend to eat a lot more than Brazilians do, but this is why the United States is the fattest country in the world.

36. Zookeepers spend more time cleaning the baboon cages than the chimpanzee cages, but this is due to the difference in behavior between both species.

37. Dolphins echolocate even more accurately than bats can, but both are amazing creatures.

38. More people gamble on March Madness than other sporting events, but many people just make brackets without any financial gain.

39. Limiting your salt usage will make a more robust tomato sauce, but a potato added to the sauce can reduce the saltiness.

40. The dragons are more able to burn villages if they have eaten a properly spicy diet.

Escher Stimuli – Experiment 3

(1) More salt-miners experienced dehydration than the florist did because of the inherent properties of salt. (Weak)

More salt-miners experienced dehydration than the florists did because of the inherent properties of salt. (Strong)

(2) More biologists studied cells than the physicist did because of the demands of the field. (Weak)

More biologists studied cells than the physicists did because of the demands of the field. (Strong)

(3) More undergraduate students skipped class than the graduate student did because oversleeping is an unhealthy habit for school. (Weak)

More undergraduate students skipped class than the graduate students did because oversleeping is an unhealthy habit for school. (Strong)

(4) More knights fought dragons than the peasant did because of the massive fires engulfing the countryside. (Weak)

More knights fought dragons than the peasants did because of the massive fires engulfing the countryside. (Strong)

(5) More paupers visited France than the prince did because of the beautiful gardens and free-flowing wine. (Weak)

More paupers visited France than the princes did because of the beautiful gardens and free-flowing wine. (Strong)

(6) More Russians caught spies than the German did because of the need to limit foreign intelligence. (Weak)

More Russians caught spies than the Germans did because of the need to limit foreign intelligence. (Strong)

(7) More ranchers herded cattle than the merchant did because of the profit obtainable from veal. (Weak)

More ranchers herded cattle than the merchant did because of the profit obtainable from veal. (Strong)

(8) More sorcerers edited tomes than the peasant did because knowledge was tightly concentrated in the kingdom. (Weak)

More sorcerers edited tomes than the peasants did because knowledge was tightly concentrated in the kingdom. (Strong)

(9) More princesses kissed frogs than the knight did because fairy tales are really weird that way. (Weak)

More princesses kissed frogs than the knights did because fairy tales are really weird that way. (Strong)

(10) More spiders trapped flies than bats did because flies are not the brightest creatures. (Weak)

More spiders trapped flies than the bat did because flies are not the brightest creatures. (Strong)

(11) More witches cut victims than the vampire did because ritual magic requires very little blood. (Weak)

More witches cut victims than the vampires did because ritual magic requires very little blood. (Strong)

(12) More programmers altered code than the marketer did because the product needed to be finished. (Weak)

More programmers altered code than the marketers did because the product needed to be finished. (Strong)

(13) More explorers travelled to new lands than the layabout did because being well-travelled has that effect.

More explorers travelled to new lands than the layabouts did because being well-travelled has that effect.

(14) More nurses helped patients than the doctor did because the hospital did not promote good bedside manner. (Weak)

More nurses helped patients than the doctors did because the hospital did not promote good bedside manner. (Strong)

(15) More line cooks left early than the sous chef did because of the stress in the kitchen. (Weak)

More line cooks left early than the sous chefs did because of the stress in the kitchen. (Strong)

(16) More cats ate dinner than the dog did because the caretaker was not very good at his job. (Weak)

More cats ate dinner than the dogs did because the caretaker was not very good at his job. (Strong)

(17) More lions played with the zookeeper than the polar bear did because of the conditions at the zoo. (Weak)

More lions played with the zookeeper than the polar bears did because of the conditions at the zoo. (Strong)

(18) More frogs hid under the lampshade than the spider did because of the warmth from the lamp. (Weak)

More frogs hid under the lampshade than the spiders did because of the warmth from the lamp. (Strong)

(19) More musicians bought candles than the artist did because of the desire to experience new things. (Weak)

More musicians bought candles than the artists did because of the desire to experience new things. (Strong)

(20) More wizards taught magic than the sorcerer did because of the dangerous curriculum that was provided. (Weak)

More wizards taught magic than the sorcerers did because of the dangerous curriculum that was provided. (Strong)

(21) More syntacticians preached Chomsky than the phonologist because of the emphasis on teaching certain theories. (Weak)

More syntacticians preached Chomsky than the phonologists because of the emphasis on teaching certain theories. (Strong)

(22) More astronomers used telescopes than the astrologist did because of the requirements of observation in science. (Weak)

More astronomers used telescopes than the astrologists did because of the requirements of observation in science. (Strong)

(23) More jocks argued over who won the football game than the nerd did because football really matters to some people. (Weak)

More jocks argued over who won the football game than the nerds did because football really matters to some people. (Strong)

(24) More snakes bit patrons than the lizard did because the zookeeper was terrible at his job. (Weak)

More snakes bit patrons than the lizards did because the zookeeper was terrible at his job. (Strong)

(25) More scientists read books than the construction worker did because of the access to information. (Weak)

More scientists read books than the construction workers did because of the access to information. (Strong)

(26) More linguistics students went to the library than the psych student did because linguistics is a very demanding field. (Weak)

More linguistics students went to the library than the psych students did because linguistics is a very demanding field. (Strong)

(27) More engineers toiled over the project than the manager did because programming can be very difficult. (Weak)

More engineers toiled over the project than the managers did because programming can be very difficult. (Strong)

(28) More soldiers experienced depression than the civilian did because war is a terrible thing. (Weak)

More soldiers experienced depression than the civilians did because war is a terrible thing. (Strong)

(29) More knights sought queens than the squire did because of the demands of the king. (Weak)

More knights sought queens than the squires did because of the demands of the king. (Strong)

(30) More nerds beat the game than the jock did because of the nature of the competition. (Weak)

More nerds beat the game than the jocks did because of the nature of the competition. (Strong)

(31) More skiers tore muscles than the ice skater did because of the danger of ski jumping. (Weak)

More skiers tore muscles than the ice skaters did because of the danger of ski jumping. (Strong)

(32) More TV anchors vomited than the cameraman did because of the anxiety of facing the public. (Weak)

More TV anchors vomited than the cameramen did because of the anxiety of facing the public. (Strong)

(33) More physics professors published articles than the literature professor did because of the nature of the profession. (Weak)

More physics professors published articles than the literature professors did because of the nature of the profession. (Strong)

(34) More women hosted potlucks than the man did because of socially constructed gender roles. (Weak)

More women hosted potlucks than the men did because of socially constructed gender roles. (Strong)

(35) More boys played hooky than the girl did because of the draconian school rules. (Weak)

More boys played hooky than the girls did because of the draconian school rules. (Strong)

(36) More kings imprisoned traitors than the duke did because of the flaws in the judicial system. (Weak)

More kings imprisoned traitors than the dukes did because of the flaws in the judicial system. (Strong)

(37) More lawyers worked overtime than the doctor did because of the extra work that comes with the job. (Weak)

More lawyers worked overtime than the doctors did because of the extra work that comes with the job. (Strong)

(38) More priests married couples than the rabbi did because of the relative size of the congregations. (Weak)

More priests married couples than the rabbis did because of the relative size of the congregations. (Strong)

(39) More bakers visited Paris than the butcher did because of the fame of French bakeries. (Weak)

More bakers visited Paris than the butchers did because of the fame of French bakeries. (Strong)

(40) More undergrads read novels than the graduate student did because of the different workloads and interests. (Weak)

More undergrads read novels than the graduate students did because of the different workloads and interests. (Strong)

(41) More NSA agents read classified files than the FBI agent did because of the war on terror. (Weak)

More NSA agents read classified files than the FBI agents did because of the war on terror. (Strong)

(42) More Venezuelans applied for visas than the Puerto Rican did because of the political unrest prevalent at the time. (Weak)

More Venezuelans applied for visas than the Puerto Ricans did because of the political unrest prevalent at the time. (Strong)

(43) More soccer fans rioted in the streets than the baseball fan did because of the world cup victory. (Weak)

More soccer fans rioted in the streets than the baseball fans did because of the world cup victory. (Strong)

(44) More backpackers visited the Grand Canyon than the skier did because of the dangerous trails and beautiful scenery. (Weak)

More backpackers visited the Grand Canyon than the skiers did because of the dangerous trails and beautiful scenery. (Strong)

(45) More princesses poisoned enemies than the witch did because of the turmoil in the kingdom. (Weak)

More princesses poisoned enemies than the witches did because of the turmoil in the kingdom. (Strong)

(46) More football players suffered from concussions than the figure skater did because of the dangers of certain sports. (Weak)

More football players suffered from concussions than the figure skaters did because of the dangers of certain sports. (Strong)

(47) More lions attacked criminals than the gladiator did because the people found animal combat entertaining. (Weak)

More lions attacked criminals than the gladiators did because the people found animal combat entertaining. (Strong)

(48) More archaeologists discovered rare fossils than the grave robber did because of the different methods of excavation. (Weak)

More archaeologists discovered rare fossils than the grave robbers did because of the different methods of excavation. (Strong)

(49) More neurosurgeons performed brain surgery than the psychiatrist did because of the surgical knowledge required. (Weak)

More neurosurgeons performed brain surgery than the psychiatrists did because of the surgical knowledge required. (Strong)

(50) More dictators ordered executions than the councilman did because of the power granted by dictatorship. (Weak)

More dictators ordered executions than the councilmen did because of the power granted by dictatorship. (Strong)

(51) More ice fishermen suffered from frostbite than the bird watcher did because of the inherent nature of cold weather sports. (Weak)

More ice fishermen suffered from frostbite than the bird watchers did because of the inherent nature of cold weather sports. (Strong)

(52) More English chefs avoided making dessert than the French chef did because of a lack of baking techniques. (Weak)

More English chefs avoided making dessert than the French chefs did because of a lack of baking techniques. (Strong)

(53) More actresses signed autographs than the dancer did because the film industry brings fame to individuals. (Weak)

More actresses signed autographs than the dancers did because the film industry brings fame to individuals. (Strong)

(54) More miners dug up diamonds than the foreman did because of the division of work responsibilities. (Weak)

More miners dug up diamonds than the foremen did because of the division of work responsibilities. (Strong)

(55) More seagulls ate trash than the pigeon did because of the nearby city population. (Weak)

More seagulls ate trash than the pigeons did because of the nearby city population. (Strong)

(56) More tourists took photos than the local did because of the interesting and unique surroundings. (Weak)

More tourists took photos than the locals did because of the interesting and unique surroundings. (Strong)

(57) More interns bought coffee than the employee did because of a need to please the office. (Weak)

More interns bought coffee than the employees did because of a need to please the office. (Strong)

(58) More carpenters hit their thumb on the job than the plumber did because hammers and saws are inherently dangerous tools. (Weak)

More carpenters hit their thumb on the job than the plumbers did because hammers and saws are inherently dangerous tools. (Strong)

(59) More park rangers rescued stranded hikers than the park visitor did because of the dangerous conditions at the park grounds. (Weak)

More park rangers rescued stranded hikers than the park visitors did because of the dangerous conditions at the park grounds. (Strong)

(60) More Russians enlisted in the Army than the American did because of the differing political and economic climates. (Weak)

More Russians enlisted in the Army than the American did because of the differing political and economic climates. (Strong)

Fillers

1.	More Michigan pet owners have cats than dogs because of the very cold weather.
2.	I think there are fewer snowstorms this winter due to global warming.
3.	Some of my friends really like electronic music, but I think classical is better.
4.	The club scene is better for dancing than bars, but I still enjoy both equally.
5.	There are a lot of animals at the zoo, but I prefer going to the aquarium.
6.	PCs can do a lot more than Macs, however most students seem to have Apple products.
7.	It has snowed more outside than any year that I can think of, and it is quite depressing.
8.	The salesman sold me way more add-ons to my cellphone than I had anticipated.

9.   Some species of insects eat their mates more often than others.
10.   Polar bears tend to have larger bodies than their cousin the panda bear.
11.   I prefer to watch Archer and Southpark over live action comedies.
12.   There are more and more chores that have to get done around the house as the weeks go on.
13.   I have spent more time on Reddit than I care to admit, but I can't stop myself.
14.   My friends all want to play video games, but I would rather read a book before bed.
15.   More students are reading spark notes these days than actually reading their books.
16.   The grocery store is only half a mile away, but I still prefer to drive to carry my groceries.
17.   My dog thinks the couch is his bed, but I prefer that he sleeps on the dog bed I bought for him.
18.   Garrett thinks that the sky is bluer in winter months than summer months.
19.   More and more snowmen are popping up all over campus due to the excess of snow.
20.   The sharpie pens are best for writing, though I sometimes prefer to use a BIC instead.
21.   Many Michiganders prefer hockey to baseball because the long, cold winters are more suited to playing hockey.
22.   Although Susie likes to cook, she prefers to bake and will spend weekends making bread.
23.   Navy lieutenants have more flying experience than Air Force lieutenants do because of the way drone warfare has increased in popularity.
24.   Because more undergrads live on campus than off campus, they become more familiar with their environment.
25.   Fewer Koreans immigrated to the United States than Irish did because Korea did not have a potato famine.
26.   Colorado has more sunny days than Michigan does, so the cold, snowy winters are beautiful.
27.   Le Anne spends more time lesson planning than Dan does, but that is because she teaches more classes.
28.   Several graduate students like to go out on Monday nights because they spend more time working over the weekend.
29.   More New Yorkers walk or take the subway than drive, but even so, traffic is terrible in New York City.
30.   Even though Iowa is a smaller state, Iowans grow more soybeans than Texans do.
31.   More Norwegians know how to ski than Italians do, and Norwegian winters are colder and longer.
32.   Fewer ranchers keep horses than farmers do because ranching requires heavy machinery and cattle.
33.   In the United States, southerners seem friendlier than northerners, but it is easier to get to know northerners.
34.   More Russians play in the NHL than Americans because Russians are better hockey players.
35.   Californians eat more avocados than Pennsylvanians do because avocados grow in California.
36.   Lawyers spend more time arguing with each other than doctors do; it's a professional hazard.

37.   In the United States, more girls play soccer than boys do, thus the women's national team is better than the men's.
38.   Fewer literature departments have internships than engineering departments do because people in the humanities like to work alone.
39.   Fewer child psychologists catch cold in the winter than pediatricians do because pediatricians work in hospitals.
40.   Eastern states have fewer water restrictions than western states do because the west is one giant desert.
41.   John thinks that his cat is prettier than Marian's dog, but their neighbors won't weigh in their opinion.
42.   I think my boyfriend has better drawing abilities than I have, but I still can cook better.
43.   There are more lions at the zoo than zebras because no one really cares about zebras.
44.   Gamers often wonder if they play more games than they are supposed to, but they don't really care.
45.   Scientists have performed many more calculations than the interns have, but they also have years of experience.
46.   There are less trains these days, but many people still find them enjoyable to ride.
47.   More students are dropping out than in previous years because our education system is growing worse and worse.
48.   There were more guys than girls at the dance but no one seemed to mind.
49.   There are more deaths from lightning than shark attacks each year, and many people are surprised at this fact.
50.   Fish tend to taste better when they come from salt water, but this is purely speculation.
51.   More cats were adopted than dogs because the shelter put the cats on a special sale.
52.   More kids ate spaghetti than tofu because they had no interest in health food.
53.   More Pharaohs were buried with piles of gold than foreign kings because gold was as common as sand in ancient Egypt.
54.   More trees were planted than cut down because of a growing environmental awareness amongst young adults.
55.   More homes are built out of brick than straw because people have learned from the experience of the three little pigs.
56.   More lamplighters burnt their fingers than candlestick makers in Victorian London because electric lights had not been invented yet.
57.   More people slipped on the sidewalks today than yesterday because of the ice storm that occurred overnight.
58.   More people watch TV than read a book because nobody likes the idea of encountering a bookworm.
59.   More people vacation in Hawaii during winter than in Michigan because people like warm weather, sunny beaches, palm trees, and hula dancers.
60.   More snow falls in Buffalo New York than Syracuse New York because of the prevalence of lake effect snow in up state New York.
61.   More nuns say their prayers everyday than other people because of the students they work with.
62.   More people eat meat than tofu because vegan diets have not caught on yet.

63.   Mercedes are more of a status symbol than Chevrolets because of the reputation of German engineers.
64.   Dogs are more useful than cats because dogs can be trained as service animals.
65.   Cathode ray tube TVs are more likely to be found in a garbage dump than on store shelves because they are an obsolete technology.
66.   Tender Vittles has become a more popular cat food than 9 lives since Morris the cat died.
67.   Broken bones are more often seen in active kids than kids sitting around playing video games.
68.   Secretaries pay more in taxes than Wall Street bankers do because the secretary's lobby is much weaker.
69.   Students skip the last class before spring break more often when they have somewhere nice to go.
70.   Kids like Christmas more than Easter because no one ever gets Easter presents.
71.   Uncle Sam wants you more than your last date because it just was not your night.
72.   There are more PCs than Macs in the computer office at work.
73.   The Easter Bunny is more popular than the health food guru because kids like candy, but only the Easter bunny like carrot sticks.
74.   Students are more likely to stay up late cramming than getting a good nights rest because they think it will actually help them perform better on the test.
75.   Horses run faster than cats because of the increased leg span and muscle mass.
76.   The girl scouts baked more cookies than the boy scouts because of the traditional values and well-renowned cookies.
77.   Rabbits move faster than turtles because of their different defense mechanisms.
78.   Lions are more ferocious than giraffes because lions must hunt other animals.
79.   Chronic back pain afflicts the elderly more than children because of the stress of living a long life.
80.   Elephants eat more than dogs because elephants are much larger creatures.
81.   Telemarketers talk more than accountants because of the requirements of their jobs.
82.   Journalists write less than editors because journalists don't pay attention to details.
83.   Magicians tend to wow audiences more than lecturers do, but I don't have a magician as a teacher.
84.   Vampires suck more blood than ticks because ticks require very little sustenance.
85.   Robots use energy more efficiently than humans because they don't create waste.
86.   Tailors sew more clothes than housewives because their income depends on it.
87.   Couch potatoes watch more TV than exercise enthusiasts because of the drastic difference in choice of hobbies.
88.   Bars sell more alcohol than restaurants because bars base their entire business model on the sale of alcohol.
89.   Cats sleep more than humans because of the different biology.
90.   Termites eat more furniture than dogs because termites eat furniture in order to survive.
91.   Libraries tend to have a wider selection of books than the bookstore does, but I can't keep the books from the library forever.
92.   Valedictorians earned higher grades than their peers because of their increased commitment to school.

93. Pirates wear more eye patches than bureaucrats do because piracy is a dangerous trade.

94. Planes travel faster than trains because planes are able to travel along a much simpler route.

95. Less time is spent smelling roses than there used to be and that is a crying shame.

96. Kids tend to eat too many sweets before bed, and this is not great for their health.

97. Far fewer people live on this planet than spiders and that terrifies me most nights.

98. More radiation is being leaked through the ozone every day because of global warming and fuel emissions.

99. The more people you have working on a project, the more headaches the manager tends to suffer.

100. More butter is called for than is necessary on boxes of macaroni and cheese because it makes everything creamier.

101. More corn will be sown than wheat next year since there will be a greater demand in the upcoming summer.

102. More shiny objects can be found in Rivendell than in Gollum's cave because Gollum is messy and no one likes him.

103. There are very few pumpkins this year compared to the last few and I have no idea why.

104. More complaints have been presented about veal than any other type of meat because of a South Park episode.

105. The frogs by the riverside tend to be happier than the neighboring frogs that live by the pond.

106. More fish swim in large ponds than lakes because ponds have fewer predators.

107. More books exist than you will ever have time to read and many will come long after you leave.

108. More quaint little towns rely on tourism to bolster their economy every year.

109. Fewer people find the time to pursue hobbies as they get older because of the demands of adulthood.

110. A greater number of Persians died at Thermopylae than was expected by any general, and it had long reaching consequences.

111. Longer surfboards make going to the beach difficult because everything has to be carried.

112. Many more accountants lived in the city than any of the surrounding suburbs and none of them liked it.

113. More computers are made every year, and they are always faster than before.

114. I need to buy more produce when I go to the grocery store, as my cart tends to fill up with junk food instead.

115. There tend to be more Android phones than iPhones that I notice people carrying around.

116. More quesadillas were ruined on the grill than burgers because tortillas are a bit more delicate than beef.

117. Fewer people these days read for pleasure, and that is really a shame considering how healthy reading is for you.

118. I have printed more papers for my anatomy class than I care to talk about.

119. Ryan counted more sheep as he attempted to fall asleep, but his insomnia was not helping.
120. More computers are broken every year due to simple mishandling and a lack of understanding viruses.

Escher Stimuli – Experiment 4

(1) More knights saw the queen than squires did in the last month. (Control)

More knights saw the queen than the squire did in the last month. (Weak)

More knights saw the queen than the squires did in the last month. (Strong)

(2) More cats ate tuna than dogs did at the pet day care center. (Control)

More cats ate tuna than the dog did at the pet day care center. (Weak)

More cats ate tuna than the dogs did at the pet day care center. (Strong)

(3) More Americans went to Russia than Brazilians did according to recent statistics. (Control)

More Americans went to Russia than the Brazilian did according to recent statistics. (Weak)

More Americans went to Russia than the Brazilians did according to recent statistics. (Strong)

(4) More forum users altered code than employees did due to recent pushes for crowd sourcing. (Control)

More forum users altered code than the employee did due to recent pushes for crowd sourcing. (Weak)

More forum users altered code than the employees did due to recent pushes for crowd sourcing. (Strong)

(5) More nurses assisted patients than doctors did according to hospital records. (Control)

More nurses assisted patients than the doctor did according to hospital records. (Weak)

More nurses assisted patients than the doctors did according to hospital records. (Strong)

(6) More lions overate than bears did because zookeepers are inconsistent. (Control)

More lions overate than the bear did because zookeepers are inconsistent. (Weak)

More lions overate than the bears did because zookeepers are inconsistent. (Strong)

(7) More frogs jumped out of the box than toads did in Will's experiment. (Control)

More frogs jumped out of the box than the toad did in Will's experiment. (Weak)

More frogs jumped out of the box than the toads did in Will's experiment. (Strong)

(8) More plumbers dropped wrenches than electricians did because of slippery conditions. (Control)

More plumbers dropped wrenches than the electrician did because of slippery conditions. (Weak)

More plumbers dropped wrenches than the electricians did because of slippery conditions. (Strong)

(9) More young adults placed bets than employees did at the local casino. (Control)

More young adults placed bets than the employee did at the local casino. (Weak)

More young adults placed bets than the employees did at the local casino. (Strong)

(10) More skiers suffered from sports related injuries than skaters did during the Olympic Games. (Control)

More skiers suffered from sports related injuries than the skater did during the Olympic Games. (Weak)

More skiers suffered from sports related injuries than the skaters did during the Olympic Games. (Strong)

(11) More interns hosted potluck dinners than employees did at the company. (Control)

More interns hosted potluck dinners than the employee did at the company. (Weak)

More interns hosted potluck dinners than the employees did at the company. (Strong)

(12) More physicists published articles than biologists did in the past three years. (Control)

More physicists published articles than the biologist did in the past three years. (Weak)

More physicists published articles than the biologists did in the past three years. (Strong)

(13) More nerds played Pokemon Go than jocks did at the local high school. (Control)

More nerds played Pokemon Go than the jock did at the local high school. (Strong)

More nerds played Pokemon Go than the jocks did at the local high school. (Weak)

(14) More writers filed lawsuits than editors did in the past decade. (Control)

More writers filed lawsuits than the editor did in the past decade. (Weak)

More writers filed lawsuits than the editors did in the past decade. (Strong)

(15) More sorcerers casted spells than wizards did due to new magical restrictions. (Control)

More sorcerers casted spells than the wizard did due to new magical restrictions. (Weak)

More sorcerers casted spells than the wizards did due to new magical restrictions. (Strong)

(16) More wizards enchanted weapons than sorcerers did due to the new laws from the king. (Control)

More wizards enchanted weapons than the sorcerer did due to the new laws from the king. (Weak)

More wizards enchanted weapons than the sorcerers did due to the new laws from the king. (Strong)

(17) More princes courted partners than princesses did in the new kingdom. (Control)

More princes courted partners than the princess did in the new kingdom. (Weak)

More princes courted partners than the princesses did in the new kingdom. (Strong)

(18) More pirates drank at the bar than rangers did in the coastal town. (Control)

More pirates drank at the bar than the ranger did in the coastal town. (Weak)

More pirates drank at the bar than the rangers did in the coastal town. (Strong)

(19) More kings imprisoned traitors than dukes did due to separate duties. (Control)

More kings imprisoned traitors than the duke did due to separate duties. (Weak)

More kings imprisoned traitors than the dukes did due to separate duties. (Strong)

(20) More doctors worked overtime than lawyers did according to records last year. (Control)

More doctors worked overtime than the lawyer did according to records last year. (Weak)

More doctors worked overtime than the lawyers did according to records last year. (Strong)

(21) More bakers visited Paris than butchers did according to internet polls. (Control)

More bakers visited Paris than the butcher did according to internet polls. (Weak)

More bakers visited Paris than the butchers did according to internet polls. (Strong)

(22) More goblins ate stolen livestock than trolls did, or so the farmer has complained. (Control)

More goblins ate stolen livestock than the troll did, or so the farmer has complained. (Weak)

More goblins ate stolen livestock than the trolls did, or so the farmer has complained. (Strong)

(23) More ghouls spooked villagers than zombies did in the town of Orion. (Control)

More ghouls spooked villagers than the zombie did in the town of Orion. (Weak)

More ghouls spooked villagers than the zombies did in the town of Orion. (Strong)

(24) More squires cleaned stables than knights did in the castle grounds. (Control)

More squires cleaned stables than the knight did in the castle grounds. (Weak)

More squires cleaned stables than the knights did in the castle grounds. (Strong)

(25) More dukes changed laws than duchesses did due to sexist law making. (Control)

More dukes changed laws than the duchess did due to sexist law making. (Weak)

More dukes changed laws than the duchesses did due to sexist law making. (Strong)

(26) More millennials sent tweets than politicians did in the last two years. (Control)

More millennials sent tweets than the politicians did in the last two years. (Weak)

More millennials sent tweets than the politician did in the last two years. (Strong)

(27) More interns bought coffee than employees did in the last fiscal year. (Control)

More interns bought coffee than the employee did in the last fiscal year. (Weak)

More interns bought coffee than the employees did in the last fiscal year. (Strong)

(28) More actresses signed autographs than actors did at this red carpet event. (Control)

More actresses signed autographs than the actor did at this red carpet event. (Weak)

More actresses signed autographs than the actors did at this red carpet event. (Strong)

(29) More ice fisherman suffered from frostbite than birdwatchers did this winter season. (Control)

More ice fisherman suffered from frostbite than the birdwatcher did this winter season. (Weak)

More ice fisherman suffered from frostbite than the birdwatchers did this winter season. (Strong)

(30) More football players suffered from concussions than snowboarders did this recent sports season. (Control)

More football players suffered from concussions than the snowboarder did this recent sports season. (Weak)

More football players suffered from concussions than the snowboarders did this recent sports season. (Strong)

Fillers – Bad

1. He said the cow more than cat the cried wolf yesterday morning.
2. Franklin compared what the cow more needed if what he asked was okay.
3. I never really liked idea of the lesser idea.
4. More people have been Russia than Brazil in the past time.
5. Modern keeping up the day with more than before I have ever seen.
6. Climbing rocks more aptly puts monks edge away in the night.
7. Will seems rampant on the wall less and less and less.
8. Aptly seeking more coats puts a janitors on some list thing.
9. Ronald ran exploding ranged fewer to the extent he could.
10. Who needs less is more than he what or she asks for.
11. Complementing fewer fines a figured one or two or more.
12. Keeping up with the times vital in the upcoming first a running.
13. New publications crumbling red will invite less guests to the party this weekend.
14. We ourselves pride at Michigan State to many more times before.
15. The dragon ate berries fewer than lesser beings in month of the may.
16. The average lifespan of the average less housefly more in the fewer month.
17. I have got to got the add the equation to the other side of the equation.
18. People more often float on ice than the sugar in the soda pop does not.
19. Really slowly move sloth the on the slow day of the month.
20. It can't ever be really the one more or less of the month is true.
21. We can take time our this month fewer than we normally could usually.
22. I remember cannot the colors of the months in the more of January.
23. Reconcile the usage of the keyboard rendering familiar faces this weekend less.
24. Forthcoming first the second of the third person to ask about this.
25. A hamburger for a coin in time the past can be your destiny one said once.
26. Promiscuous promising prove will will not abide by the several counting.
27. Captain experience red seas soaring under the ocean the green.
28. A way to aubergine open more fluidly involves a little roast.
29. Make them dog gold it up more and less these days.
30. What to hold is nearly the greenery beckoning the forthcoming time apart.
31. Randy needs adjustment an and more or less is what we want from that.
32. That was a greatly annulment from raven the the that people tells us all about.
33. See here is hearing what we hear ports directly into greater amounts.
34. A cat can't see red the reading room with the tickling mices.
35. If you dry the river cannot greater befalls the one who does not cannot.

36. Writing more and more will inevitably prove degrees than greater than can you dream of it.
37. I had my understandings at the time lessening the lessons I heard never.
38. Running in circles ringing the door random puts on alert the great movement.
39. More company picnics go awry alrightly in midnight sun summer vacation.
40. You sometimes have make to the choice to go or me, that's it.
41. Waiting on the doorbell watched pot never boils as my greater grandmother once said.
42. Greener greenery can off putting be from the eye of beholder the says he.
43. But I cannot tell you cleaner stories from dragging the dock outside.
44. Sung tales seems sillier than seemed said so what.
45. If you put the key down glossier cabinet the dropped we can get a new one later.
46. Come forth redder to places understanding that you will become the greater red.
47. Putting clothes too in the washer many can yield blended colors you don't want.
48. It's both the dreams granted fashioned less people dancing around the clock all night.
49. Music to ears my more on the radio than I had anticipated earlier today.
50. Wolverines are scarier creatures I had imagined than than before.
51. The shrimp pasties delicious are they than grosser.
52. Reading too many video games arguably television many more.
53. Less knowledged folks the figures show are in need of too.
54. Definitely not what magic expected a greater showing that before that.
55. Porcupines frequently unearth cats the cats the neighbor more than likely hates.
56. Exploring the deep inside ocean submarine requires greater risk than usual.
57. A highway travels cars down less country roads that travel do.
58. I can't understand cannot but notice the similarities between the sets.
59. On a whim dogs the sniff fire hydrants on passersby walking.
60. Basketball can played more with seven participants on the field.

Fillers – Good

1. More Michigan pet owners have cats than dogs because of the very cold weather.
2. I think there are fewer snowstorms this winter due to global warming.
3. Some of my friends really like electronic music, but I think classical is better.
4. The club scene is better for dancing than bars, but I still enjoy both equally.
5. There are a lot of animals at the zoo, but I prefer going to the aquarium.
6. PCs can do a lot more than Macs, however most students seem to have Apple products.
7. It has snowed more outside than any year that I can think of, and it is quite depressing.
8. The salesman sold me way more add-ons to my cellphone than I had anticipated.
9. Some species of insects eat their mates more often than others.
10. Polar bears tend to have larger bodies than their cousin the panda bear.
11. I prefer to watch Archer and Southpark over live action comedies.
12. There are more and more chores that have to get done around the house as the weeks go on.

13. I have spent more time on Reddit than I care to admit, but I can't stop myself.

14. My friends all want to play video games, but I would rather read a book before bed.

15. More students are reading spark notes these days than actually reading their books.

16. The grocery store is only half a mile away, but I still prefer to drive to carry my groceries.

17. My dog thinks the couch is his bed, but I prefer that he sleeps on the dog bed.

18. More and more snowmen are popping up all over campus due to the excess of snow.

19. The sharpie pens are best for writing, though I sometimes prefer to use a BIC instead.

20. Many Michiganders prefer hockey to baseball because the long, cold winters are more suited to playing hockey.

21. Although Susie likes to cook, she prefers to bake and will spend weekends making bread.

22. I like milkshakes more than malts, but that is a personal preference.

23. Because more undergrads live on campus than off campus, they become more familiar with their environment.

24. Colorado has more sunny days than Michigan does, so the cold, snowy winters are beautiful.

25. More New Yorkers walk or take the subway than drive, but even so, traffic is terrible in New York City.

26. Even though Iowa is a smaller state, Iowans grow more soybeans than Texans do.

27. More Norwegians know how to ski than Italians do, and Norwegian winters are colder and longer.

28. Fewer ranchers keep horses than farmers do because ranching requires heavy machinery and cattle.

29. In the United States, southerners seem friendlier than northerners, but it is easier to get to know northerners.

30. Californians eat more avocados than Pennsylvanians do because avocados grow in California.

31. Lawyers spend more time arguing with each other than doctors do; it's a professional hazard.

32. Eastern states have fewer water restrictions than western states do because the west is one giant desert.

33. John thinks that his cat is prettier than Marian's dog, but their neighbors won't weigh in their opinion.

34. There are more lions at the zoo than zebras because no one really cares about zebras.

35. Gamers often wonder if they play more games than they are supposed to, but they don't really care.

36. Scientists have performed more calculations than the interns have, but they also have years of experience.

37. There are fewer trains these days, but many people still find them enjoyable to ride.

38. More students are dropping out than in previous years because our education system is growing worse and worse.

39. There were more guys than girls at the dance but no one seemed to mind.

40. There are more deaths from lightning than shark attacks each year, and many people are surprised at this fact.

41. Fish tend to taste better when they come from salt water, but this is purely speculation.

42. More cats were adopted than dogs because the shelter put the cats on a special sale.

43. More trees were planted than cut down because of a growing environmental awareness.

44. More homes are built of brick than of wood due to the availability of certain materials.

45. More people slipped on the sidewalks today than yesterday because of the ice storm that occurred overnight.

46. More people watch TV than read a book because nobody likes the idea of encountering a bookworm.

47. More snow falls in Buffalo New York than Syracuse New York because of lake effect snow.

48. More nuns say their prayers everyday than other people because of the students they work with.

49. More people eat meat than tofu because vegan diets have not caught on yet.

50. Cathode ray tube TVs are more likely to be found in a garbage dump than on store shelves.

51. Students skip the last class before spring break more often when they have somewhere nice to go.

52. There are more PCs than Macs in the computer office at work.

53. Journalists write less than editors because journalists don't pay attention to details.

54. Robots use energy more efficiently than humans because they don't create waste.

55. Tailors sew more clothes than housewives because their income depends on it.

56. More radiation is being leaked through the ozone every day because of global warming and fuel emissions.

57. The more people you have working on a project, the more headaches the manager tends to suffer.

58. The frogs by the riverside tend to be happier than the neighboring frogs that live by the pond.

59. More quaint little towns rely on tourism to bolster their economy every year.

60. More computers are made every year, and they are always faster than before.

Escher Stimuli – Experiment 5

(1) More ogres ate trash than the donkeys did. (Escher)

More ogres ate trash than donkeys did. (Control)

More ogres ate trash than the donkeys did which made the swamp a little cleaner. (Extended)

More ogres ate trash than the donkeys ate trash which made the swamp a little cleaner. (Non Ellipsis)

(2) More cows gave milk than the goats did. (Escher)

More cows gave milk than goats did. (Control)

More cows gave milk than the goats did which did not shock the farmer. (Extended)

More cows gave milk than the goats gave milk which did not shock the farmer. (Non Ellipsis)

(3) More computer scientists coded software than the interns did. (Escher)

More computer scientists coded software than interns did. (Control)

More computer scientists coded software than the interns did which upset the manager. (Extended)

More computer scientists coded software than the interns coded software which upset the manager. (Non Ellipsis)

(4) More government workers propagated lies than the politicians did. (Escher)

More government workers propagated lies than politicians did. (Control)

More government workers propagated lies than the politicians did which surprised everyone. (Extended)

More government workers propagated lies than the politicians propagated lies to which surprised everyone. (Non Ellipsis)

(5) More butchers carved turkeys than the grocers did. (Escher)

More butchers carved turkeys than grocers did. (Control)

More butchers carved turkeys than the grocers did for all the necessary Thanksgiving preparations. (Extended)

More butchers carved turkeys than the grocers carved turkeys for all the necessary Thanksgiving preparations. (Non Ellipsis)

(6) More flies flew around garbage than the crickets did. (Escher)

More flies flew around garbage than crickets did. (Control)

More flies flew around garbage than the crickets did even though there was no rotting food. (Extended)

More flies flew around garbage than the crickets flew around the garbage even though there was no rotting food. (Non Ellipsis)

(7) More owls ate mice than the eagles did. (Escher)

More owls ate mice than eagles did. (Control)

More owls ate mice than the eagles did, which terrified the rodents. (Extended)

More owls ate mice than the eagles ate mice, which terrified the rodents. (Non Ellipsis)

(8)More dolphins played with bubbles than the children did. (Escher)

More dolphins played with bubbles than children did. (Control)

More dolphins played with bubbles than the children did, and everyone had fun at the aquarium. (Extended)

More dolphins played with bubbles than the children played with bubbles, and everyone had fun at the aquarium. (Non Ellipsis)

(9)More rats looked for cheese than the mice did. (Escher)

More rats looked for cheese than mice did. (Control)

More rats looked for cheese than the mice did, which upset the scientist's hypothesis. (Extended)

More rats looked for cheese than the mice looked for cheese, which upset the scientist's hypothesis. (Non Ellipsis)

(10) More undergrads took notes in class than the grad students did. (Escher)

More undergrads took notes in class than grad students did. (Control)

More undergrads took notes in class than the grad students did, and the test results reflected that. (Extended)

More undergrads took notes in class than the grad students took notes in class, and the test results reflected that. (Non Ellipsis)

(11) More farmers shopped online than the scientists did. (Escher)

More farmers shopped online than scientists did. (Control)

More farmers shopped online than the scientists did even though access to computers was difficult. (Extended)

More farmers shopped online than the scientists shopped online even though access to computers was difficult. (Non Ellipsis)

(12) More ghosts haunted villages than the zombies did. (Escher)

More ghosts haunted villages than zombies did. (Control)

More ghosts haunted villages than the zombies did this past harvest season. (Extended)

More ghosts haunted villages than the zombies haunted villages this past harvest season. (Non Ellipsis)

(13) More knights slew dragons than the wizards did. (Escher)

More knights slew dragons than wizards did. (Control)

More knights slew dragons than the wizards did during the great sundering. (Extended)

More knights slew dragons than the wizards slew dragons during the great sundering. (Non Ellipsis)

(14) More gravediggers used cash for gold than the surfers did. (Escher)

More gravediggers used cash for gold than surfers did. (Control)

More gravediggers used cash for gold than the surfers did due to the nature of their jobs. (Extended)

More gravediggers used cash for gold than the surfers used cash for gold due to the nature of their jobs. (Non Ellipsis)

(15) More police ate hamburgers than the firefighters did. (Escher)

More police ate hamburgers than firefighters did. (Control)

More police ate hamburgers than the firefighters did, but the sewage workers ate more. (Extended)

More police ate hamburgers than the firefighters ate hamburgers, but the sewage workers ate more. (Non Ellipsis)

(16) More grad students played D&D than the undergrads did. (Escher)

More grad students played D&D than undergrads did. (Control)

More grad students played D&D than the undergrads did because most students are sick of video games. (Extended)

More grad students played D&D than the undergrads played D&D because most students are sick of video games. (Non Ellipsis)

(17) More scholars read scrolls than the barbarians did. (Escher)

More scholars read scrolls than barbarians did. (Control)

More scholars read scrolls than the barbarians did, which makes a lot of sense. (Extended)

More scholars read scrolls than the barbarians read scrolls, which makes a lot of sense. (Non Ellipsis)

(18) More blacksmiths smoked tobacco than the weavers did. (Escher)

More blacksmiths smoked tobacco than weavers did. (Control)

More blacksmiths smoked tobacco than the weavers did due to the smoking nature of the guild. (Extended)

More blacksmiths smoked tobacco than the weavers smoked tobacco due to the smoking nature of the guild. (Non Ellipsis)

(19) More undergrads procrastinated than the grad students did. (Escher)

More undergrads procrastinated than grad students did. (Control)

More undergrads procrastinated than the grad students did, but everyone had some stress. (Extended)

More undergrads procrastinated than the grad students procrastinated, but everyone had some stress. (Non Ellipsis)

(20) More grad students panicked at exam time than the undergrads did. (Escher)

More grad students panicked at exam time than undergrads did. (Control)

More grad students panicked at exam time than the undergrads did because of everyone's study habits. (Extended)

More grad students panicked at exam time than the undergrads panicked at exam time because of everyone's study habits. (Non Ellipsis)

(21) More customers dropped food than the cooks did. (Escher)

More customers dropped food than cooks did. (Control)

More customers dropped food than the cooks did which upset the host. (Extended)

More customers dropped food than the cooks dropped food which upset the host. (Non Ellipsis)

(22) More employees stole merchandise than the managers did. (Escher)

More employees stole merchandise than managers did. (Control)

More employees stole merchandise than the managers did which was heavily admonished by corporate. (Extended)

More employees stole merchandise than the managers stole merchandise which was heavily admonished by corporate. (Non Ellipsis)

(23) More airlines gauged their prices than the retailers did. (Escher)

More airlines gauged their prices than retailers did. (Control)

More airlines gauged their prices than the retailers did during the past holiday sale. (Extended)

More airlines gauged their prices than the retailers gauged their prices during the past holiday sale. (Non Ellipsis)

(24) More cats sat in boxes than the dogs did. (Escher)

More cats sat in boxes than dogs did. (Control)

More cats sat in boxes than the dogs did when I worked at the pet day care center. (Extended)

More cats sat in boxes than the dogs sat in boxes when I worked at the pet day care center. (Non Ellipsis)

(25) More dogs wanted to play than the cats did. (Escher)

More dogs wanted to play than cats did. (Control)

More dogs wanted to play than the cats did when I was watching my neighbor's pets. (Extended)

More dogs wanted to play than the cats wanted to play when I was watching my neighbor's pets. (Non Ellipsis)

(26) More tigers slept at the zoo than the pandas did. (Escher)

More tigers slept at the zoo than pandas did. (Control)

More tigers slept at the zoo than the pandas did which made for an unexciting day. (Extended)

More tigers slept at the zoo than the pandas slept at the zoo which made for an unexciting day. (Non Ellipsis)

(27) More reindeer played games than the elves did. (Escher)

More reindeer played games than elves did. (Control)

More reindeer played games than the elves did, but all had fun this holiday season. (Extended)

More reindeer played games than the elves played games, but all had fun this past holiday season. (Non Ellipsis)

(28) More giants traversed mountains than the hobbits did. (Escher)

More giants traversed mountains than hobbits did. (Control)

More giants traversed mountains than the hobbits did, but there were tumultuous storms. (Extended)

More giants traversed mountains than the hobbits traversed mountains, but there were tumultuous storms. (Non Ellipsis)

(29) More lawyers went out to lunch than the assistants did. (Escher)

More lawyers went out to lunch than assistants did. (Control)

More lawyers went out to lunch than the assistants did because some like to talk business then. (Extended)

More lawyers went out to lunch than the assistants went out to lunch because some like to talk business then. (Non Ellipsis)

(30) More doctors stayed up late than the nurses did. (Escher)

More doctors stayed up late than nurses did. (Control)

More doctors stayed up late than the nurses did because of new hospital policies. (Extended)

More doctors stayed up late than the nurses stayed up late because of new hospital policies. (Non Ellipsis)

(31) More birds ate seed than the squirrels did. (Escher)

More birds ate seed than squirrels did. (Control)

More birds ate seed than the squirrels did due to the number of feeders. (Extended)

More birds ate seed than the squirrels ate seed due to the number of feeders. (Non Ellipsis)

(32) More frogs jumped across the road than the raccoons did. (Escher)

More frogs jumped across the road than raccoons did. (Control)

More frogs jumped across the road than the raccoons did, and traffic came to a standstill. (Extended)

More frogs jumped across the road than the raccoons jumped across the road, and traffic came to a standstill. (Non Ellipsis)

(33) More knights fought invading dragons than the peasants did. (Escher)

More knights fought invading dragons than peasants did. (Control)

More knights fought invading dragons than the peasants did, and the villagers were pleased. (Extended)

More knights fought invading dragons than the peasants fought invading dragons, and the villagers were pleased. (Non Ellipsis)

(34) More wizards casted spells than the sorcerers did. (Escher)

More wizards casted spells than sorcerers did. (Control)

More wizards casted spells than the sorcerers did which left the magicians in a quandary. (Extended)

More wizards casted spells than the sorcerers casted spells which left the magicians in a quandary. (Non Ellipsis)

(35) More rafters packed energy bars than the hikers did. (Escher)

More rafters packed energy bars than hikers did. (Control)

More rafters packed energy bars than the hikers did because of the needs of the race. (Extended)

More rafters packed energy bars than the hikers packed energy bars because of the needs of the race. (Non Ellipsis)

(36) More zombies devoured villagers than the goblins did. (Escher)

More zombies devoured villagers than goblins did. (Control)

More zombies devoured villagers than the goblins did, and the king was not so happy. (Extended)

More zombies devoured villagers than the goblins devoured villagers, and the king was not so happy. (Non Ellipsis)

(37) More beauticians donated money than the barbers did. (Escher)

More beauticians donated money than barbers did. (Control)

More beauticians donated money than the barbers did because some people are stingy. (Extended)

More beauticians donated money than the barbers donated money because some people are stingy. (Non Ellipsis)

(38) More Brits drank tea at the party than the Americans did. (Escher)

More Brits drank tea at the party than Americans did. (Control)

More Brits drank tea at the party than the Americans did which surprised no one. (Extended)

More Brits drank tea at the party than the Americans drank tea which surprised no one. (Non Ellipsis)

(39) More baboons played with toys than the chimpanzees did. (Escher)

More baboons played with toys than chimpanzees did. (Control)

More baboons played with toys than the chimpanzees did, and the zookeepers were perplexed. (Extended)

More baboons played with toys than the chimpanzees did, and the zookeepers were perplexed. (Non Ellipsis)

(40) More grandmothers baked cookies than the mothers did. (Escher)

More grandmothers baked cookies than mothers did. (Control)

More grandmothers baked cookies than the mothers did, but the kids didn't care. (Extended)

More grandmothers baked cookies than the mothers baked cookies, but the kids didn't care. (Non Ellipsis)

(41) More nerds played board games than the jocks did. (Escher)

More nerds played board games than jocks did. (Control)

More nerds played board games than the jocks did, which did not surprise the cheerleaders. (Extended)

More nerds played board games than the jocks played board games, which did not surprise the cheerleaders. (Non Ellipsis)

(42) More republicans ate pizza than the democrats did. (Escher)

More republicans ate pizza than democrats did. (Control)

More republicans ate pizza than the democrats did, which was a topic at the convention. (Extended)

More republicans ate pizza than the democrats ate pizza, which was a topic at the convention. (Non Ellipsis)

(43) More Americans swam in the ocean than the Brazilians did. (Escher)

More Americans swam in the ocean than Brazilians did. (Control)

More Americans swam in the ocean than the Brazilians did, but the sharks didn't care. (Extended)

More Americans swam in the ocean than the Brazilians swam in the ocean, but the sharks didn't care. (Non Ellipsis)

(44) More Brits visited the Tower of London than the Americans did. (Escher)

More Brits visited the Tower of London than Americans did. (Control)

More Brits visited the Tower of London than the Americans did, or so they say. (Extended)

More Brits visited the Tower of London than the Americans visited the tower of London, or so they say. (Non Ellipsis)

(45) More cats have scratched the pantry doors than the dogs have. (Escher)

More cats have scratched the pantry doors than dogs have. (Control)

More cats have scratched the pantry doors than the dogs have which upset the owner. (Extended)

More cats have scratched the pantry doors than the dogs have scratched the pantry doors which upset the owner. (Non Ellipsis)

(46) More catfish swallowed hooks than the trouts did. (Escher)

More catfish swallowed hooks than trouts did. (Control)

More catfish swallowed hooks than the trouts did, which surprised the fisherman. (Extended)

More catfish swallowed hooks than the trouts swallowed hooks, which surprised the fisherman. (Non Ellipsis)

(47) More towels were washed than the sheets were. (Escher)

More towels were washed than sheets were. (Control)

More towels were washed than the sheets were, but there was too much laundry. (Extended)

More towels were washed than the sheets were washed, but there was too much laundry. (Non Ellipsis)

(48) More bats flew through the yard than the birds did. (Escher)

More bats flew through the yard than birds did. (Control)

More bats flew through the yard than the birds did, which upset the neighborhood. (Extended)

More bats flew through the yard than the birds flew through the yard, which upset the neighborhood. (Non Ellipsis)

(49) More students drank coffee outside than the faculty did. (Escher)

More students drank coffee outside than faculty did. (Control)

More students drank coffee outside than the faculty did, but everyone still had pumpkin spice. (Extended)

More students drank coffee outside than the faculty drank coffee outside, but everyone still had pumpkin spice. (Non Ellipsis)

(50) More carpenters shopped at home depot than the plumbers did. (Escher)

More carpenters shopped at home depot than plumbers did. (Control)

More carpenters shopped at home depot than the plumbers did, but everyone else chose Menards. (Extended)

More carpenters shopped at home depot than the plumbers shopped at home depot, but everyone else chose Menards. (Non Ellipsis)

(51) More locals skateboarded at the park than the tourists did. (Escher)

More locals skateboarded at the park than tourists did. (Control)

More locals skateboarded at the park than the tourists did, but there were still rad moves. (Extended)

More locals skateboarded at the park than the tourists skateboarded at the park, but there were still rad moves. (Non Ellipsis)

(52) More squirrels ate nuts than the raccoons did. (Escher)

More squirrels ate nuts than raccoons did. (Control)

More squirrels ate nuts than the raccoons did, but the chipmunks ate the most. (Extended)

More squirrels ate nuts than the raccoons ate nuts, but the chipmunks ate the most. (Non Ellipsis)

(53) More hippies played hackey sack than the frat guys did. (Escher)

More hippies played hackey sack than frat guys did. (Control)

More hippies played hackey sack than the frat guys did, but fun was had by all. (Extended)

More hippies played hackey sack than the frat guys played hackey sack, but fun was had by all. (Non Ellipsis)

(54) More rhinos drank from the hole than the ostriches did. (Escher)

More rhinos drank from the hole than ostriches did. (Control)

More rhinos drank from the hole than the ostriches did, though the elephants drank the most. (Extended)

More rhinos drank from the hole than the ostriches drank from the hole, though the elephants drank the most. (Non Ellipsis)

(55) More magicians tricked people than the insurance agents did. (Escher)

More magicians tricked people than insurance agents did. (Control)

More magicians tricked people than the insurance agents did, given it was a magic convention. (Extended)

More magicians tricked people than the insurance agents tricked people, given it was a magic convention. (Non Ellipsis)

(56) More pigs rolled in mud than the dogs did. (Escher)

More pigs rolled in mud than dogs did. (Control)

More pigs rolled in mud than the dogs did, and the humans were thankful. (Extended)

More pigs rolled in mud than the dogs rolled in the mud, and the humans were thankful. (Non Ellipsis)

(57) More Brazilians made sandwiches than the Americans did. (Escher)

More Brazilians made sandwiches than Americans did. (Control)

More Brazilians made sandwiches than the Americans did, yet they were all delicious. (Extended)

More Brazilians made sandwiches than the Americans made sandwiches, yet they were all delicious. (Non Ellipsis)

(58) More cats caused mischief than the dogs did. (Escher)

More cats caused mischief than dogs did. (Control)

More cats caused mischief than the dogs did, but the birds were well behaved. (Extended)

More cats caused mischief than the dogs caused mischief, but the birds were well behaved. (Non Ellipsis)

(59) More turtles swam in the river than the frogs did. (Escher)

More turtles swam in the river than frogs did. (Control)

More turtles swam in the river than the frogs did, and the herons were just as happy. (Extended)

More turtles swam in the river than the frogs swam in the river, and the herons were just as happy. (Non Ellipsis)

(60) More bikers ran the red light than the cab drivers did. (Escher)

More bikers ran the red light than cab drivers did. (Control)

More bikers ran the red light than the cab drivers did, and the police kept busy. (Extended)

More bikers ran the red light than the cab drivers ran the red light, and the police kept busy. (Non Ellipsis)

Fillers – Bad

1. She can't the cat see, but the dog she definitely see.
2. More rats under the table than under the bush were they.
3. Mark needed to plants the water the bathtub under the sink did she.
4. Fewer orangutans can understand can understand the understanding.
5. The wires to the computer connect cannot the to the printer.
6. No matter where green is located under the mattress, she will always bring a peach.
7. Contacting the spirit realm caused Ryu under watching the TV last night.
8. Lip service for the rich ran through default text even once.
9. More linguists linguisted linguistics linguistcally languished.
10. I'm a rebel even with the forty six under cats the.
11. I remember when a famous singer song wrote the best lyrics ever.
12. You can say one thing mean but another to another one another.

13. One may wonder whether more oranges help can the duty officer from the park yesterday.

14. She chased the running walls until dawn the of new year.

15. Now now said now what now is now until knowledge abound.

16. Miranda wondered if the cats outside were left outside in the outside outside.

17. I can make a mistake, but I can't the mistake make me.

18. More apples the in fall fall falling through the window.

19. It is just way the of the world that I can't have superpowers from the full foom.

20. Sometimes time flies by without even mention a the project due at midnight.

21. Will considers that the snacks provided may the force with him to be.

22. The man was coming out cage the place he did not want to go.

23. Listening to music distracting during phases of the initiation.

24. Captioned the captain before he casts the aside the boat.

25. My preference for video games over books is something me dearly cost in me grades.

26. Ryan that believes information together put can be to be memorized more easily.

27. The dodo don't exist anymore, and and consequence the consequence is minor.

28. Wells difficult are dig to in the wintertime because of the frozen ground.

29. Looking at the harvest, farmer the that decided to close down the farm.

30. Creating the complex rigorous work from the scientists involved not.

31. Maybe Jason rediscover will understand the final answer to it.

32. Exporting the fermenting process pleasing will he now get the picture.

33. Polly you stop the police internet from getting rebooted stairs under there.

34. More the I think about the answer, more the I don't even.

35. Computers can calculate numbers randomly the the scientists understand do not.

36. Every star has it's share fair atoms of in the nucleotides it seems to have.

37. Anywhere is applicable the to the nordic underpinnings of the computer.

38. Quentin looked at alley the the bar behind and noticed all the discareded televisions.

39. Fewer than the average bear flew around moon the a couple of grand times.

40. Czech buildings have advantage architecture allimony alligators.

41. More than nothing something the an expert can probably tell instead.

42. Expressions are hard compute to to put lightly in the toaster for the best settings.

43. Fulfilling the request Joe believes best to Japan back to fly anyway.

44. Squaring your shoulders for a good game golf proves for vital for for success.

45. Needing to fold more laundry, the socks buttered toast the in the pantry.

46. Yes or no responses last year so are last year was definitely not at to the least last year.

47. The most exciting times are nose running under the right nose.

48. Wordsworth remembered that the the understatement of the century needed amplification.

49. What is surreal is what is real the in the understanding of the scientists.

50. Paul error more than usual last year, and it is probably error due to the exam.

51. More cakes need baking according to to fulfill before the request by the chef.

52. The drummer broke to sticks the on his drumset the night before.

53. Guitarists are more likely to more than the drummer not can understand.

54. Knights are less likely to scold the cleric party in the adventurer's guild.

55. Rae wishes her dog talk could could understand her that much better.
56. Yardwork can be a pain, but payoffs the the greatness of it is definitely worth it.
57. Sooner better the building the lego structure the quicker it can be put away.
58. Socks tend to get lost in the dry, and who knows go where they go happens that when.
59. Julia needed a new headset, so she more tried to buy less money the store.
60. Calling Arthur requires digits the from the secret notepad that man the passed to me at the restaurant.

Fillers – Good

1.   More and more I like dogs more than I like cats.
2.   More people need surgery for gangrene these days.
3.   More pieces of furniture are in this house than I can handle.
4.   More owls are smarter than the average bear.
5.   More people believe that coral reefs are prettier than evergreen forests, but I disagree.
6.   More people believe there is more evidence for a flat earth surfacing every day.
7.   More time studying is typically needed as you get better and better at it.
8.   More folk think that Shelia believes less fluoride is being added to her water supply than usual.
9.   More students believe that photographers are more pretentious than artists typically are.
10.   More industries every year wants to make us think that there are less pollutants entering the water.
11.   More of his friends think John likes Mary more than he likes Randy.
12.   More of his friends think Randy does not care how much more John likes him than Mary.
13.   More owls make worse pets than parakeets do.
14.   More pineapples were sold than apples this year.
15.   More and more video games are added to my library every day.
16.   Fewer than 100 bengal tigers exist in the wild today.
17.   I like to eat burritos more than tacos, but both are pretty good.
18.   Ryan believes there are less people involved in politics today.
19.   There are way too many lightswitches in this room.
20.   Sticky notes are less useful than moleskin notepads.
21.   I find more and more bugs in this room than I care to admit.
22.   Telephones are getting to be cheaper and cheaper as time goes on.
23.   Kaylin needs more plates in her apartment.
24.   Fewer owners are feeding their dogs store bought food these days.
25.   The trash heap is taller than my roommate, and he is seven feet tall.
26.   Will wonders if there are more snacks inside the fridge than yesterday.
27.   T Swift has more and more concerts every year.
28.   I can't imagine that I could eat any more food.
29.   You would think less people would consider the flat earth theory is real.
30.   There are more fruit in this basket than I had anticipated.

31.   There was a lot of soda served at the party yesterday.
32.   No one has ever seen big foot in real life, though many believe he or she exists.
33.   The Loch Ness Monster probably laughs at all the bad photos of it.
34.   Phone companies have tried to charge us for every little feature.
35.   The panels are not the right color, or so the designer says.
36.   Yoshi wanted to buy more ice cream, but the store ran out.
37.   Maryanne was upset about her roses in the garden.
38.   Drinking diet soda is probably worse for you than regular soda, but I am no doctor.
39.   Robert believes his computer is out to get him, but maybe google is.
40.   I dropped my keys down the sewer, but that's a no go after watching It.
41.   The socket requires extra fittings to work properly.
42.   Erin wanted to fix her pond, but the bottom tarp was ripped.
43.   Peter spoke at the conference yesterday, but he forgot his slides.
44.   I think Phyllis wants to go on another vacation this summer.
45.   Bill had a hard time keeping up with the payments on his new motorcycle.
46.   Lee thinks that his friends are a bunch of nerds.
47.   Jason wants to believe that he can play basketball, but he needs to work on his jumps.
48.   Sean can always bring the life to the party.
49.   James needed a haircut, so he visited his local barber.
50.   Some will wonder why Matthew works for that company.
51.   People think that Jacob is pretentious, but I think that he just needs to work on his attitude.
52.   I was so busy thinking about what I could do that I forgot to do it.
53.   Stephanie is always in sync with everyone's emotions.
54.   Emily wasted a ton of paper printing out her dissertation, but the office didn't mind.
55.   No one can convince Carol that she needs to brush her teeth.
56.   Raul remembers that he has to set the oven to 450 before his partner gets home.
57.   Danielle needs to replace that old chair in the living room because it is unsafe to sit on.
58.   I can't tell if my new phone is working or not.
59.   There are several movies that most people consider to be the best ever.
60.   Robin wondered what the meaning of life is.

Experiment 6 Stimuli

(1) More ogres ate trash than donkeys ate trash. (NE Control)

More ogres ate trash than donkeys did. (Control)

More ogres ate trash than the donkeys did which made the swamp a little cleaner. (Extended)

More ogres ate trash than the donkeys ate trash which made the swamp a little cleaner. (Non Ellipsis)

(2) More cows gave milk than goats gave milk. (NE Control)

More cows gave milk than goats did. (Control)

More cows gave milk than the goats did which did not shock the farmer. (Extended)

More cows gave milk than the goats gave milk which did not shock the farmer. (Non Ellipsis)

(3) More computer scientists coded software than interns coded software. (NE Control)

More computer scientists coded software than interns did. (Control)

More computer scientists coded software than the interns did which upset the manager. (Extended)

More computer scientists coded software than the interns coded software which upset the manager. (Non Ellipsis)

(4) More government workers propagated lies than politicians propagated lies. (NE Control)

More government workers propagated lies than politicians did. (Control)

More government workers propagated lies than the politicians did which surprised everyone. (Extended)

More government workers propagated lies than the politicians propagated lies to which surprised everyone. (Non Ellipsis)

(5) More butchers carved turkeys than grocers carved turkeys. (NE Control)

More butchers carved turkeys than grocers did. (Control)

More butchers carved turkeys than the grocers did for all the necessary Thanksgiving preparations. (Extended)

More butchers carved turkeys than the grocers carved turkeys for all the necessary Thanksgiving preparations. (Non Ellipsis)

(6) More flies flew around garbage than crickets flew around garbage. (NE Control)

More flies flew around garbage than crickets did. (Control)

More flies flew around garbage than the crickets did even though there was no rotting food. (Extended)

More flies flew around garbage than the crickets flew around the garbage even though there was no rotting food. (Non Ellipsis)

(7) More owls ate mice than eagles ate mice. (NE Control)

More owls ate mice than eagles did. (Control)

More owls ate mice than the eagles did, which terrified the rodents. (Extended)

More owls ate mice than the eagles ate mice, which terrified the rodents. (Non Ellipsis)

(8) More dolphins played with bubbles than children played with bubbles. (NE Control)

More dolphins played with bubbles than children did. (Control)

More dolphins played with bubbles than the children did, and everyone had fun at the aquarium. (Extended)

More dolphins played with bubbles than the children played with bubbles, and everyone had fun at the aquarium. (Non Ellipsis)

(9) More rats looked for cheese than mice looked for cheese. (NE Control)

More rats looked for cheese than mice did. (Control)

More rats looked for cheese than the mice did, which upset the scientist's hypothesis. (Extended)

More rats looked for cheese than the mice looked for cheese, which upset the scientist's hypothesis. (Non Ellipsis)

(10) More undergrads took notes in class than grad students took notes in class. (NE Control)

More undergrads took notes in class than grad students did. (Control)

More undergrads took notes in class than the grad students did, and the test results reflected that. (Extended)

More undergrads took notes in class than the grad students took notes in class, and the test results reflected that. (Non Ellipsis)

(11) More farmers shopped online than scientists shopped online. (NE Control)

More farmers shopped online than scientists did. (Control)

More farmers shopped online than the scientists did even though access to computers was difficult. (Extended)

More farmers shopped online than the scientists shopped online even though access to computers was difficult. (Non Ellipsis)

(12) More ghosts haunted villages than zombies haunted villages. (NE Control)

More ghosts haunted villages than zombies did. (Control)

More ghosts haunted villages than the zombies did this past harvest season. (Extended)

More ghosts haunted villages than the zombies haunted villages this past harvest season. (Non Ellipsis)

(13) More knights slew dragons than wizards slew dragons. (NE Control)

More knights slew dragons than wizards did. (Control)

More knights slew dragons than the wizards did during the great sundering. (Extended)

More knights slew dragons than the wizards slew dragons during the great sundering. (Non Ellipsis)

(14) More gravediggers used cash for gold than surfers used cash for gold. (NE Control)

More gravediggers used cash for gold than surfers did. (Control)

More gravediggers used cash for gold than the surfers did due to the nature of their jobs. (Extended)

More gravediggers used cash for gold than the surfers used cash for gold due to the nature of their jobs. (Non Ellipsis)

(15) More police ate hamburgers than firefighters ate hamburgers. (NE Control)

More police ate hamburgers than firefighters did. (Control)

More police ate hamburgers than the firefighters did, but the sewage workers ate more. (Extended)

More police ate hamburgers than the firefighters ate hamburgers, but the sewage workers ate more. (Non Ellipsis)

(16) More grad students played D&D than undergrads played D&D. (NE Control)

More grad students played D&D than undergrads did. (Control)

More grad students played D&D than the undergrads did because most students are sick of video games. (Extended)

More grad students played D&D than the undergrads played D&D because most students are sick of video games. (Non Ellipsis)

(17) More scholars read scrolls than barbarians read scrolls. (NE Control)

More scholars read scrolls than barbarians did. (Control)

More scholars read scrolls than the barbarians did, which makes a lot of sense. (Extended)

More scholars read scrolls than the barbarians read scrolls, which makes a lot of sense. (Non Ellipsis)

(18) More blacksmiths smoked tobacco than weavers smoked tobacco. (NE Control)

More blacksmiths smoked tobacco than weavers did. (Control)

More blacksmiths smoked tobacco than the weavers did due to the smoking nature of the guild. (Extended)

More blacksmiths smoked tobacco than the weavers smoked tobacco due to the smoking nature of the guild. (Non Ellipsis)

(19) More undergrads procrastinated than grad students procrastinated. (NE Control)

More undergrads procrastinated than grad students did. (Control)

More undergrads procrastinated than the grad students did, but everyone had some stress. (Extended)

More undergrads procrastinated than the grad students procrastinated, but everyone had some stress. (Non Ellipsis)

(20) More grad students panicked at exam time than undergrads panicked at exam time. (NE Control)

More grad students panicked at exam time than undergrads did. (Control)

More grad students panicked at exam time than the undergrads did because of everyone's study habits. (Extended)

More grad students panicked at exam time than the undergrads panicked at exam time because of everyone's study habits. (Non Ellipsis)

(21) More customers dropped food than cooks dropped food. (NE Control)

More customers dropped food than cooks did. (Control)

More customers dropped food than the cooks did which upset the host. (Extended)

More customers dropped food than the cooks dropped food which upset the host. (Non Ellipsis)

(22) More employees stole merchandise than managers stole merchandise. (NE Control)

More employees stole merchandise than managers did. (Control)

More employees stole merchandise than the managers did which was heavily admonished by corporate. (Extended)

More employees stole merchandise than the managers stole merchandise which was heavily admonished by corporate. (Non Ellipsis)

(23) More airlines gauged their prices than retailers gauged their prices. (NE Control)

More airlines gauged their prices than retailers did. (Control)

More airlines gauged their prices than the retailers did during the past holiday sale. (Extended)

More airlines gauged their prices than the retailers gauged their prices during the past holiday sale. (Non Ellipsis)

(24) More cats sat in boxes than dogs sat in boxes. (NE Control)

More cats sat in boxes than dogs did. (Control)

More cats sat in boxes than the dogs did when I worked at the pet day care center. (Extended)

More cats sat in boxes than the dogs sat in boxes when I worked at the pet day care center. (Non Ellipsis)

(25) More dogs wanted to play than cats wanted to play. (NE Control)

More dogs wanted to play than cats did. (Control)

More dogs wanted to play than the cats did when I was watching my neighbor's pets. (Extended)

More dogs wanted to play than the cats wanted to play when I was watching my neighbor's pets. (Non Ellipsis)

(26) More tigers slept at the zoo than pandas slept at the zoo. (NE Control)

More tigers slept at the zoo than pandas did. (Control)

More tigers slept at the zoo than the pandas did which made for an unexciting day. (Extended)

More tigers slept at the zoo than the pandas slept at the zoo which made for an unexciting day. (Non Ellipsis)

(27) More reindeer played games than elves played games. (NE Control)

More reindeer played games than elves did. (Control)

More reindeer played games than the elves did, but all had fun this holiday season. (Extended)

More reindeer played games than the elves played games, but all had fun this past holiday season. (Non Ellipsis)

(28) More giants traversed mountains than hobbits travesed mountains. (NE Control)

More giants traversed mountains than hobbits did. (Control)

More giants traversed mountains than the hobbits did, but there were tumultuous storms. (Extended)

More giants traversed mountains than the hobbits traversed mountains, but there were tumultuous storms. (Non Ellipsis)

(29) More lawyers went out to lunch than assistants went out to lunch. (NE Control)

More lawyers went out to lunch than assistants did. (Control)

More lawyers went out to lunch than the assistants did because some like to talk business then. (Extended)

More lawyers went out to lunch than the assistants went out to lunch because some like to talk business then. (Non Ellipsis)

(30) More doctors stayed up late than nurses stayed up late. (NE Control)

More doctors stayed up late than nurses did. (Control)

More doctors stayed up late than the nurses did because of new hospital policies. (Extended)

More doctors stayed up late than the nurses stayed up late because of new hospital policies. (Non Ellipsis)

(31) More birds ate seed than squirrels ate seed. (NE Control)

More birds ate seed than squirrels did. (Control)

More birds ate seed than the squirrels did due to the number of feeders. (Extended)

More birds ate seed than the squirrels ate seed due to the number of feeders. (Non Ellipsis)

(32) More frogs jumped across the road than raccoons jumped across the road. (NE Control)

More frogs jumped across the road than raccoons did. (Control)

More frogs jumped across the road than the raccoons did, and traffic came to a standstill. (Extended)

More frogs jumped across the road than the raccoons jumped across the road, and traffic came to a standstill. (Non Ellipsis)

(33) More knights fought invading dragons than peasants fought invading dragons. (NE Control)

More knights fought invading dragons than peasants did. (Control)

More knights fought invading dragons than the peasants did, and the villagers were pleased. (Extended)

More knights fought invading dragons than the peasants fought invading dragons, and the villagers were pleased. (Non Ellipsis)

(34) More wizards casted spells than sorcerers casted spells. (NE Control)

More wizards casted spells than sorcerers did. (Control)

More wizards casted spells than the sorcerers did which left the magicians in a quandary. (Extended)

More wizards casted spells than the sorcerers casted spells which left the magicians in a quandary. (Non Ellipsis)

(35) More rafters packed energy bars than hikers packed energy bars. (NE Control)

More rafters packed energy bars than hikers did. (Control)

More rafters packed energy bars than the hikers did because of the needs of the race. (Extended)

More rafters packed energy bars than the hikers packed energy bars because of the needs of the race. (Non Ellipsis)

(36) More zombies devoured villagers than goblins devoured villagers. (NE Control)

More zombies devoured villagers than goblins did. (Control)

More zombies devoured villagers than the goblins did, and the king was not so happy. (Extended)

More zombies devoured villagers than the goblins devoured villagers, and the king was not so happy. (Non Ellipsis)

(37) More beauticians donated money than barbers donated money. (NE Control)

More beauticians donated money than barbers did. (Control)

More beauticians donated money than the barbers did because some people are stingy. (Extended)

More beauticians donated money than the barbers donated money because some people are stingy. (Non Ellipsis)

(38) More Brits drank tea at the party than Americans drank tea at the party. (NE Control)

More Brits drank tea at the party than Americans did. (Control)

More Brits drank tea at the party than the Americans did which surprised no one. (Extended)

More Brits drank tea at the party than the Americans drank tea which surprised no one. (Non Ellipsis)

(39) More baboons played with toys than chimpanzees played with toys. (NE Control)

More baboons played with toys than chimpanzees did. (Control)

More baboons played with toys than the chimpanzees did, and the zookeepers were perplexed. (Extended)

More baboons played with toys than the chimpanzees did, and the zookeepers were perplexed. (Non Ellipsis)

(40) More grandmothers baked cookies than mothers baked cookies. (NE Control)

More grandmothers baked cookies than mothers did. (Control)

More grandmothers baked cookies than the mothers did, but the kids didn't care. (Extended)

More grandmothers baked cookies than the mothers baked cookies, but the kids didn't care. (Non Ellipsis)

Fillers – Bad

1. She can't the cat see, but the dog she definitely see.
2. More rats under the table than under the bush were they.
3. Mark needed to plants the water the bathtub under the sink did she.
4. Fewer orangutans can understand can understand the understanding.
5. The wires to the computer connect cannot the to the printer.
6. No matter where green is located under the mattress, she will always bring a peach.
7. Contacting the spirit realm caused Ryu under watching the TV last night.
8. Lip service for the rich ran through default text even once.
9. More linguists linguisted linguistics linguistcally languished.
10. I'm a rebel even with the forty six under cats the.
11. I remember when a famous singer song wrote the best lyrics ever.

12. You can say one thing mean but another to another one another.
13. One may wonder whether more oranges help can the duty officer from the park yesterday.
14. She chased the running walls until dawn the of new year.
15. Now now said now what now is now until knowledge abound.
16. Miranda wondered if the cats outside were left outside in the outside outside.
17. I can make a mistake, but I can't the mistake make me.
18. More apples the in fall fall falling through the window.
19. It is just way the of the world that I can't have superpowers from the full foom.
20. Sometimes time flies by without even mention a the project due at midnight.
21. Will considers that the snacks provided may the force with him to be.
22. The man was coming out cage the place he did not want to go.
23. Listening to music distracting during phases of the initiation.
24. Captioned the captain before he casts the aside the boat.
25. My preference for video games over books is something me dearly cost in me grades.
26. Ryan that believes information together put can be to be memorized more easily.
27. The dodo don't exist anymore, and and consequence the consequence is minor.
28. Wells difficult are dig to in the wintertime because of the frozen ground.
29. Looking at the harvest, farmer the that decided to close down the farm.
30. Creating the complex rigorous work from the scientists involved not.
31. Maybe Jason rediscover will understand the final answer to it.
32. Exporting the fermenting process pleasing will he now get the picture.
33. Polly you stop the police internet from getting rebooted stairs under there.
34. More the I think about the answer, more the I don't even.
35. Computers can calculate numbers randomly the the scientists understand do not.
36. Every star has it's share fair atoms of in the nucleotides it seems to have.
37. Anywhere is applicable the to the nordic underpinnings of the computer.
38. Quentin looked at alley the the bar behind and noticed all the discareded televisions.
39. Fewer than the average bear flew around moon the a couple of grand times.
40. Czech buildings have advantage architecture allimony alligators.

Fillers – Good

1.   More and more I like dogs more than I like cats.
2.   More people need surgery for gangrene these days.
3.   More pieces of furniture are in this house than I can handle.
4.   More owls are smarter than the average bear.
5.   More people believe that coral reefs are prettier than evergreen forests, but I disagree.
6.   More people believe there is more evidence for a flat earth surfacing every day.
7.   More time studying is typically needed as you get better and better at it.
8.   More folk think that Shelia believes less fluoride is being added to her water supply than usual.
9.   More students believe that photographers are more pretentious than artists typically are.

10. More industries every year wants to make us think that there are less pollutants entering the water.
11. Fewer than 100 bengal tigers exist in the wild today.
12. I like to eat burritos more than tacos, but both are pretty good.
13. Ryan believes there are less people involved in politics today.
14. There are way too many lightswitches in this room.
15. Sticky notes are less useful than moleskin notepads.
16. I find more and more bugs in this room than I care to admit.
17. Telephones are getting to be cheaper and cheaper as time goes on.
18. Kaylin needs more plates in her apartment.
19. Fewer owners are feeding their dogs store bought food these days.
20. The trash heap is taller than my roommate, and he is seven feet tall.
21. There was a lot of soda served at the party yesterday.
22. No one has ever seen big foot in real life, though many believe he or she exists.
23. The Loch Ness Monster probably laughs at all the bad photos of it.
24. Phone companies have tried to charge us for every little feature.
25. The panels are not the right color, or so the designer says.
26. Yoshi wanted to buy more ice cream, but the store ran out.
27. Maryanne was upset about her roses in the garden.
28. Drinking diet soda is probably worse for you than regular soda, but I am no doctor.
29. Robert believes his computer is out to get him, but maybe google is.
30. I dropped my keys down the sewer, but that's a no go after watching It.
31. The socket requires extra fittings to work properly.
32. Erin wanted to fix her pond, but the bottom tarp was ripped.
33. Peter spoke at the conference yesterday, but he forgot his slides.
34. I think Phyllis wants to go on another vacation this summer.
35. Bill had a hard time keeping up with the payments on his new motorcycle.
36. Lee thinks that his friends are a bunch of nerds.
37. Jason wants to believe that he can play basketball, but he needs to work on his jumps.
38. Sean can always bring the life to the party.
39. James needed a haircut, so he visited his local barber.
40. Some will wonder why Matthew works for that company.

REFERENCES

REFERENCES

Abney, S. P. (1987). *The English noun phrase in its sentential aspect*. Massachusetts Institute of Technology.

Bach, M., & Poloschek, C. M. (2006). Optical illusions. *Adv Clin Neurosci Rehabil*, *6*(2), 20–21.

BBC News. (2015). Optical illusion: Dress colour debate goes global.

beim Graben, P., Gerth, S., & Vasishth, S. (2008). Towards dynamical system models of language-related brain potentials. *Cognitive Neurodynamics*, *2*(3), 229–255.

Bock, K., & Miller, C. (1991). Broken Agreement. *Cognitive Psychology*, *23*, 45–93.

Bošković, Ž. (2005). On the locality of left branch extraction and the structure of NP*. *Studia Linguistica*, *59*(1), 1–45.

Bresnan, J. W. (1973). Syntax of the Comparative Clause Construction in English. *Linguistic Inquiry*, *IV*(3), 275–343.

Carlson, G. (1977). *Reference to kinds in English.* (Dissertation). University of Massachusetts, Amherst.

Chomsky, N. (1956). On the limits of finite-state description. *MIT Research Laboratory for Electronics, Quarterly Progress Report 41*, 64–65.

Chomsky, N., & Miller, G. (1963). *Introduction to the formal analysis of natural languages.* (Vol. 2). New York: Wiley.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836.

Clifton, C., & Frazier, L. (1989). Comprehending Sentences with Long-Distance Dependencies. In *Linguistic Structure in Language Processing* (pp. 273–317). Springer, Dordrecht.

Culicover, P., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Drenhaus, H., Frisch, S., & Saddy, D. (2005). Processing Negative Polarity Items: When Negation Comes Through the Backdoor. *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 145–165.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making Syntax of Sense: Number Agreement. *Psychological Review*, *112*(3), 1-29.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*(1), 3–20.

Ferreira, F., & Patson, N. D. (2007). The "Good Enough" Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83.

Fitzroy, A. B., & Sanders, L. D. (2013). Musical Expertise Modulates Early Processing of Syntactic Violations in Language. *Frontiers in Psychology*, *3*.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*(3), 427–473.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* University of Connecticut.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science*, *6*, 78–84.

Fruchter, J., & Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain and Language*, *143*, 81–96.

Fults, S., & Phillips, C. (2004). *The source of syntactic illusions.* CUNY.

Giannakidou, A. (2011). 64. Negative and positive polarity items. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science*. Berlin, Boston: De Gruyter.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments: Linguistic Acceptability on Mechanical Turk. *Language and Linguistics Compass*, *5*(8), 509–524.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248.

Gow, D. W., Keller, C. J., Eskandar, E., Meng, N., & Cash, S. S. (2009). Parallel versus serial processing dependencies in the perisylvian speech network: A Granger analysis of intracranial EEG data. *Brain and Language*, *110*(1), 43–48.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112.

Grice, H. P. (1975). Logic and Conversation. In *Syntax and Semantics, 3: Speech Acts*. Academic Press.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases.* Princeton University.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar.* Malden, MA: Blackwell Publishing.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*(3), 425–443.

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, *86*(2), 366–415.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*(2), 365–392.

Kehler, A. (2002). *Coherence in discourse.* Stanford, California: CSLI Publications.
Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225.

Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology*, *68*(6), 540–546.

Krifka, M. (1990). Four Thousand Ships Passed Through The Lock: Object-Induced Measure Functions on Events. *Linguistics and Philosophy*, *13*, 487–520.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect sematnic incongruity. *Science*, *207*, 203–208.

Kutas, M., Van Patten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II: 1994-2005. *Handbook of Psycholinguistics*, 659–724.

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by 'the dress' photograph. *Current Biology*, *25*(13), R545–R546.

Lasnik, H. (1972). *Analysis of Negation in English*. Massachusetts Institute of Technology.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933.

Lee, D., & Newman, S. D. (2009). The effect of presentation paradigm on syntactic processing: An event-related fMRI study. *Human Brain Mapping*, 1–15.

Likert, R. (1932). *Likert technique for attitude measurement*. New York University. Retrieved from Archives of Psychology, 22, 5–55.

Luck, S. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, Massachusetts: The MIT Press.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 887–904.

Mancini, S., Molinaro, N., Davidson, D. J., Avilés, A., & Carreiras, M. (2014). Person and the syntax–discourse interface: An eye-tracking study of agreement. *Journal of Memory and Language*, *76*, 141–157.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(4), 1057–1074.

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60.

Mathworks. (2018). *MATLAB*. Mathworks.

Mauchly, J. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209.

Mechanical Turk. (2017). Seattle, Washington: Amazon.

Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis.* Oxford: Oxford University Press.

Merchant, J. (2013). Voice and ellipsis. *Linguistic Inquiry*, *44*(1), 77–108.

Montalbetti, M. M. (1984). *After binding: On the interpretation of pronouns*. Dissertation. Massachusetts Institute of Technology.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*(2), 151–165.

O'Connor, E. (2015). *Comparative illusions at the syntax-semantics interface*. Dissertation. University of Southern California.

O'Connor, E., Pancheva, R., & Kaiser, E. (2012). Evidence for online repair of Escher sentences1. In *Proceedings of Sinn und Bedeutung*, 17, 363–380.

Osterhout, L., & Holcomb, P. J. (1992). Event-Related Brain Potentials Elicited by Syntactic Anomaly. *Journal of Memory and Language*, *31*, 785–806.

Peirce, J. (2018). *PsychoPy*.

Phillips, C. (2013). Parser-grammar relations: We don't understand everything twice. *Oxford University Press*, 294–315.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical Illusions and Selective Fallibility in Real-Time Language Comprehension. Runner, Jeffrey (Ed.): Experiments at the Interfaces. *Syntax and Semantics*, 37, 147–180, Emerald Group Publishing Limited.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*(1), 73–193.

Qualtrics. (2017). Provo, Utah: Qualtrics.

R Core Team. (2013). *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The Effect of Clause Wrap-Up on Eye Movements during Reading. *The Quarterly Journal of Experimental Psychology Section A*, *53*(4), 1061–1080.

Rooth, M. (1993). *Ellipsis Redundancy and Reduction Redundancy*.

Ross, J. R. (1967). *Constraints on variables in syntax.* Dissertation. MIT.

Sag, I. (1976). *Deletion and logical form.* Dissertation. MIT.

Schütze, C. T., & Sprouse, J. (2014). Judgment data. *Research Methods in Linguistics*, 27–50.
Sohn, K.-W. (1995). Negative polarity items, scope, and economy.

Solomyak, O., & Marantz, A. (2009). Evidence for Early Morphological Decomposition in Visual Word Recognition. *Journal of Cognitive Neuroscience*, *22*(9), 2042–2057.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, *43*(1), 155–167.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135–162.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. *Science*, *150*, 1187–1188.

Szabolcsi, A. (2006). Strong vs. Weak Islands. *The Blackwell Companion to Syntax*, *4*, 479–531.

Townsend, J. T. (1990). Processing: Sometimes They Look Like Tweedledum and Tweedledee but They Can (and Should) be Distinguished. *Psychological Science*, *1*(1), 46–54.

Uribe-Echevarria, M. (1994). *Interface licensing conditions on negative polarity items: a theory of polarity and tense interactions.* University of Connecticut.

Vasishth, S., Brussow, S., Lewis, R., & Drenhaus, H. (2008). Processing Polarity: How the Ungrammatical Intrudes on the Grammatical. *Cognitive Science: A Multidisciplinary Journal*, *32*(4), 685–712.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380–384.

Wellwood, A. (2015). On the semantics of comparison across categories. *Linguistics and Philosophy*, *38*(1), 67–101.

Wellwood, A., Pancheva, R., Hacquard, V., Fults, S., & Phillips, C. (2009). *The role of event comparison in comparative illusions.* Presented at the CUNY.

Wellwood, A., Pancheva, R., Hacquard, V., & Phillips, C. (2017). The anatomy of a comparative illusion. *Ms., University of Maryland, College Park*.

Whelpton, M., Trotter, D., Beck, Þ. G., Anderson, C., Maling, J., Durvasula, K., & Beretta, A. (2014). Portions and sorts in Icelandic: An ERP study. *Brain and Language*, *136*, 44–57.