"Money or River": A Bistable Approach to Investigating the Neural Correlates of Lexical Ambiguity

---

A Thesis

Presented to

The Division of The Established Interdisciplinary Committee for Neuroscience

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Declan Greenwald

May 2023

Approved for the Division

(Neuroscience)

Enriqueta Canseco-Gonzalez

Acknowledgments

Below I have taken the time and space to thank those who have helped me along this journey of college, life, and all that jazz.

Taran, my brother, your constant support throughout my life has helped shape me into the person I am today. Knowing you always have my back is a constant comfort. I love you.

Mom and Dad, OBVIOUSLY you must be thanked for your unconditional love and care. Thank you for putting up with me and my many fits and conniptions. I've always felt safe and comfortable being myself around you, a comfort and privilege not everyone is afforded. I love ya!

To my grandparents, who I really should call more often. Sorry about that. Thank you for your hand in giving me this opportunity to pursue my studies here at Reed. I love you and I hope to see you soon.

To Chris for nearly two decades of friendship. You have the remarkable ability of making many of my memories of lockdown pleasant ones. To many decades of more wonderful memories.

To Scud T. Dog, though we have known each other for fourteen years, it is almost impossible to describe how thankful I am for these last three years specifically.

To Adrian, Alex, Farhan, Maddy, and Olivia, who mean more to me they know.

To Mary Greenburg for providing me with my wardrobe

To Griffin, David, Ben, and Marie for altering my vocabulary forever.

To my roommate Ari, thank you for putting up with my shenanigans. I wish you all the best.

Enriqueta, thank you so much for your guidance and all of your help throughout this project. I wish you nothing but peace and prosperity in your retirement.

Cole, without your help I could not have completed this project. Thank you so much for teaching me everything I needed to know about the EEG process.

Thank you to Michael for all of your insights and help with this project and thank you to everyone in the SCALP lab for your help and guidance. A special thanks to everyone in the lab who helped me cap participants.

Thank you to Jennifer for being a wonderful and supportive academic advisor these last few years.

Thanks to everyone I have undoubtedly forgotten who helped me get to where I am right now

And of course, thank you to everyone who participated in this study. I'm sorry if I pulled out any of your hair.

Table of Contents

List of Tables

List of Figures

Abstract

The study of how the brain processes ambiguous visual stimuli (e.g., Boring's Old/Young Woman) has provided the fields of psychology and neuroscience with wide swaths of knowledge pertaining to how we form a semblance of coherent sensory experience from the never-ending stream of sensory stimuli we encounter in daily life. The Reversal Negativity (RN) is an event-related potential (ERP) elicited when an onlooker's subjective perception of an ambiguous bistable figure switches from one of its possible interpretations to the other. The RN is thought to reflect neural processes behind this perceptive change. In 2018, Reed College senior Kevin Ortego set out to investigate whether ambiguous sentences having two valid interpretations (e.g., "The chicken is ready to eat.") are neurally represented in a similar fashion to these ambiguous figures. Using a novel approach to a standard reversal task where participants were asked if their interpretation of the ambiguous sentence matched with a disambiguating drawing (i.e., a hungry or cooked chicken), Ortego compared ERPs elicited by these disambiguating stimuli in mismatching (i.e., reversal) reports vs. matching (i.e., stable) reports. In response to reversals of "bistable" ambiguous sentences, they identified a large, frontally distributed negativity effect occurring over a similar time-course as the visual RN, deemed the "conceptual" Reversal Negativity. The present study aims to further Ortego's research by removing a potential confound. Here we investigate whether ambiguous words having two valid interpretations (e.g., "BANK" meaning a financial institution or the land alongside a river) are represented in a similar way as ambiguous bistable figures. To investigate this question, we recorded brain activity in fourteen participants while presenting ambiguous words many times in a row. Each ambiguous word was given two potential meanings, and on each trial, participants indicated which meaning they "perceived" or conceptualized. We then compared ERPs elicited by these ambiguous stimuli in trials where the response matched the previous response (i.e., stable) vs. trials where the response differed from the previous response (i.e., reversal). We observed two early occipital negativities in reversal trials which could each be identified as an RN, or as two parts of the same reversal mechanism. We also identified what could be defined as either a Late Positive Complex or a P300(P3) in reversal trials, which is consistent with bistable figure paradigms. We interpret this finding as evidence that the brain may engage in similar types of processing and perceptual switching across different types of bistable ambiguities, in this case for single word ambiguities. We discuss potential alternative explanations for these findings.

For Kazoo

Chapter 1: Introduction

1.1 Ambiguity

Life may at times be unpredictable, but we tend to go about life with the assumption that what we see, and experience is congruent with the physical reality of the world. However, we occasionally encounter stimuli that force us to perceive things outside of their physical reality. The most easily accessible, and pop culturally ubiquitous, category of such stimuli would be optical illusions. Such illusions make us ponder how much of our perception is skewed from physical reality. Illusions offer researchers an opportunity to investigate the neural correlates that underlie the ways in which people process visual ambiguities.

Using visual illusions allows us to separate out the processes of sensation, where a sense organ (our eyes in the case of visual illusions) detects external stimuli and sends a signal to the brain, and perception, which refers to the processes our brain employs to make sense of these signals before ultimately arriving at a conscious and subjective experience of the outside world, although where sensation ends, and perception begins is not a well-defined line (Kornmeier et al. 2012). The separation of sensation and perception is seen in the kinds of optical illusions the viewer perceives something that is not actually in the stimulus itself (Bach and Poloschek 2006). Such illusions (See Fig 1.1) have traditionally been described as the result of interactions between ganglion cells in the retina, which are separated by only one layer of cells from the rods and cones responsible for a litany of the sensations that build our sense of vison. However, there are countless numbers of different intricacies in various optical illusions and many theories on the underlying electrophysiology inherent in their perception.

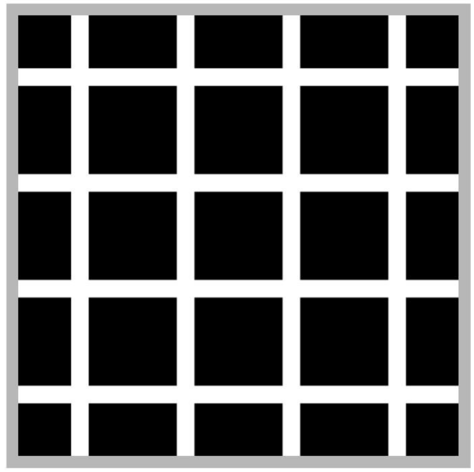

Figure 1.1 The Hermann Grid
The Hermann Grid is an example of an optical illusion where the viewer perceives something that is not physically present in the stimulus. When viewing this grid, you should see gray dots at the intersection points of the perpendicular white lines. However, when one tries to look directly at one specific gray dot, it disappears.

The reason why optical illusions are so fascinating and why I want to answer these thorny questions brings us to the topic of how to investigate perceptual ambiguity and its role in perception more broadly. Our thoughts, biased opinions, emotions, memories, mental frameworks, and experiences of shared events often differ from those around us. This leads to the question; how do we even study a fundamental and mutually objective internal illusory experience when certain objective stimuli have many simultaneous subjective yet real internal perceptions? One such way researchers investigate this topic is with bistable figures.

1.2 Bistable Figures

Bistable figures are visual stimuli which can be perceived in one of two mutually exclusive ways. Some famous examples include the Necker Cube, Boring's Old/Young Woman, Schroeder's Staircase, and the Jastow rabbit-duck

(See Fig. 1.2). Our perception of each of these images readily switches between two interpretations despite the fact that the image itself remains static. A crucial element of these figures is that only one interpretation can be perceived at a time. Even though we know that both interpretations are possible, our perception is limited to one interpretation at a time.

The reader of this thesis should spend some time looking at the figures below and personally experience the bistable switching between the two interpretations. Sometimes you can even intentionally switch your interpretation, or keep one interpretation stable for a long time, but you cannot see both interpretations at once. What I find most fascinating is doing nothing at all but staring at the middle of each figure and experiencing spontaneous switching. It is a visceral feeling when you experience this kind of shift, and it can be quite charming. The well documented and bizarre nature of these reversals in perception has made them a subject of research interest, particularly for neuroscientists investigating the link between sensation and perception. In particular, one question is what happens in the brain when these switches occur? To answer this question, the ERP technique is often used, allowing us to measure brain electrical activity time-locked to a specific event. Below in section 1.3 I describe the ERP technique in more detail, but first, I would like to briefly mention an important distinction made by a couple of researchers in regard to the particular nature of bistable images in the study of consciousness.

1.2.1 Bistable Paradigms and "Conscious Interpretation."

Recently, Bachmann & Aru (2023) describe a crucial difference in bistable paradigms in comparison to the majority of other forms of investigation into neural correlates of consciousness (NCC). As opposed to experimental approaches like binocular rivalry, which function in such a way that a target stimulus is either present in a participant's conscious experience or not, in the case of bistable stimuli, the content remains constant, and the effect monitored is a change in the *interpretation* of that content. Bachmann and Aru describe the contrast between different interpretations of conscious contents of the already consciously experienced stimulation as "conscious interpretation," a term this study will use from here on out.

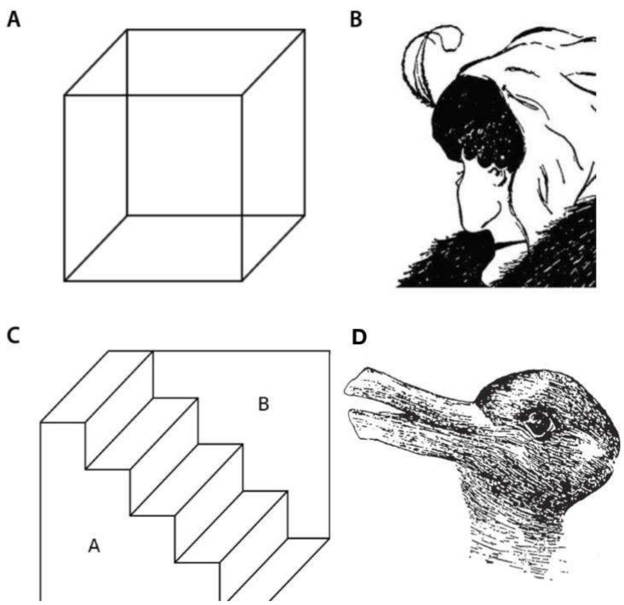

Figure 1.2 Examples of bistable figures (Ortego, 2018) (A) The Necker Cube. The lower left square can appear either in the foreground as the "front" of the cube, or in the background with the upper right square forming the "top" of the other cube interpretation. (B) A variant of Boring's Old/Young Woman. The old woman looks down and to the left. The young woman looks away over her should and wears "the old woman's mouth" as a necklace. (C) Schroeder's stairs can be seen as a normal staircase descending from the top left to the bottom right with region A in the foreground, or as a staircase with its back hanging from the "ceiling" and descending from the top right to the bottom left, with region B in the foreground. (D) The Jastow rabbit-duck. This image presents a hand-drawn image of an animal, switching between a rabbit facing right, or a duck facing left, depending on whether you interpret the left side of the image as the rabbits' ears or the duck's beak. Should any of these images remain stagnant, just stare at the center of the image and your perception will eventually switch.

1.3 EEG and ERPs

Electroencephalography (EEG) is a technique that uses an array of electrodes placed on the scalp in order to measure brain electrical activity. EEG is a measure of a multitude of simultaneous postsynaptic potentials. Postsynaptic potentials are the electrochemical result of a neurotransmitter binding to the dendrites of postsynaptic neurons. Postsynaptic potentials are generated from neurons oriented similarly in the cortical tissue, and perpendicular to the surface such that they summate to create a strongly enough field to be detected by electrodes on the scalp. This is what we record and call the electroencephalogram or EEG (Luck 2014).

The brain is constantly rife with neuronal activity, and as such, it'd be very difficult to extract any sort of a specific neuronal correlation of cognitive function. However, by recording EEG while presenting many trials and "time-locking" the resulting data to some specific event, such as the appearance of a stimulus or a button press (to indicate a response), random electrical noise tends to cancel itself out after the averaging of many trials. This means that any activity that stands out is likely related to the time-locked event. Such signals are therefore referred to as event-related potentials (ERPs). By comparing ERPs elicited under different experimental conditions, we can examine differences in brain activity on a time scale of milliseconds. Some particular features of ERP waveforms are reliably associated with a specific task or stimulus, such as hearing a deviant tone in a stream of consistent tones or deciding about whether to respond to a stimulus or not. These consistently observed features of a waveform are referred to as "ERP components" and often receive names that reflect their voltage, timescale, or function (Luck 2014).

1.4 EEG Correlates of Bistable Perception

1.4.1 Reversal Negativity

To better understand the process of a reversal I invite you to once again look at one of the bistable stimuli in figure 1.2. As previously stated, subjective interpretation of sequentially presented bistable figures will switch spontaneously if/when you look at the stimulus' center. However, this time, when looking at one of the figures, blink roughly every half second. Every few blinks, your interpretation of the figure should switch upon opening your eyes. Congratulations! You have just experienced a reversal. In a lab setting, instead of asking participants to blink, experimenters flash the bistable figure repeatedly at a semi-constant rate while participants report for each stimulus what they saw, and then afterward the experimenter determines for each trial whether there was a reversal of interpretation or not. This is what we call a "reversal task". So essentially a reversal trial is when a participant's conscious interpretation of an ambiguous stimulus changes (from the previous presentation), and a stable trial is when a participant's conscious interpretation of an ambiguous stimulus stays constant.

The Reversal Negativity (RN) is an ERP component that appears approximately 260ms after stimulus presentation associated with a wide range of ambiguous stimuli (Kornmeier and Bach 2004, 2005; Pitts, et al., 2007, 2008). It is called the RN because in bistable paradigms, reversal trials elicit a larger negativity than stable trials in certain scalp regions (predominantly in occipital and parietal areas of the scalp). The duration and topographic distribution of the RN are sensitive to a variety of factors, most notably response interval (Kornmeier & Bach, 2012), which suggests that this component can be modulated by factors beyond changes in perception. However, the extent to which non-perceptual effects do or do not interact with the RN is not yet fully understood.

The RN is calculated by taking the averaged ERPs in response to trials where a reversal of perception occurs and subtracting from that the average of ERPs of trials where perception remained stable. Plainly, something different happens in your brain when either your conscious interpretation or perception of the image changes and that difference manifests as the RN. A reversal itself is actually relatively rare (compared to stable trials), generally occurring usually around 30% of the time in a standard reversal task (Kornmeier & Bach, 2004; Pitts et al., 2007).

Researchers have investigated whether this kind of endogenous (i.e., spontaneous) switching in bistable percepts is different from exogenous switches triggered by a disambiguating physical change in the stimulus. This is normally accomplished by adding or removing a crucial feature to the image (See fig 1.3). Such research has found that the brain does respond differently to stimulus-driven exogenous reversals and percept-driven endogenous reversals. In these exogenous trials, an RN is still observed but it peaks around 220ms after stimulus onset, as compared to 260ms for endogenous reversals (Kornmeier and Bach 2006). Another ERP signature, a Reversal Positivity (RP) seen ~130ms after stimulus onset, has also been observed in response to these exogenous reversals (Kornmeier and Bach 2012). Kornmeier and Bach (2011) consider the RP to be reflective of an initial detection of stimulus ambiguity via processing conflict at the stage of 3D object interpretation prior to the perceptual reversal (as indexed by the later RN), an interpretation that fits with the fact that the RP occurs only in response to reversals of ambiguous figures and not their disambiguated (or unambiguous) variants.

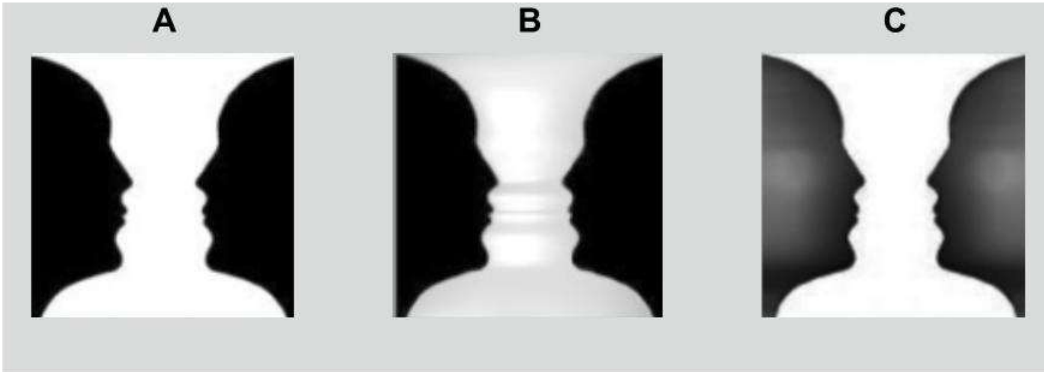

Figure 1.3 Different representations of Rubin's bistable Face-Vase figure. The ambiguous bistable face-vase (A) can be seen disambiguated with changes in shading and light to look like (B) an unambiguous vase (in white), or (C) an unambiguous pair of profiles (in black).

The functional role or cause of the RN is still a relative mystery, considering it is subject to bottom-up (i.e., stimulus driven) and top-down (i.e., interpretation, attention, etc.) influences and occurs across many different paradigms and sensory modalities. There is an argument that it may simply be indicative of a general response to the kind of change in the perceptual understanding of a representation (i.e., it represents the change of seeing the Necker cube one way to the other way) in the brain (Intaite et al. 2010, Kornmeier and Bach 2012). The amplitude of the RN has also been found to increase when participants are instructed to intentionally cause reversals, compared to reporting when reversals spontaneously occur (Pitts et al. 2008). This would suggest that top-down control modulates some of the machinations behind the processing of bistable figures very early after stimulus onset. However, bottom-up factors, such as duration of stimulus presentation and inter-stimulus intervals also affect amplitude of the RN (Kornmeier et al. 2007)

Previous research has looked into the RN's relationship to the attention related N2pc, an ERP component found in posterior scalp regions contralateral to the presented stimulus, by using a bilateral display of two bistable figures and examining whether an N2pc occurred when one of the two figures reversed

(Intaite et al. 2010). The authors suggest that the RN is *not* a variant of the N2pc, which suggest that attention is not the only driving force behind the RN. They note that the RN did not vary with the specific variants of the bistable figures used. Kornmeier and Bach (2012) offer an explanation framed in terms of switches between the multiple perceptual "attractor" states which might correspond to the current ambiguous visual stimulus. Take an unambiguous image of a cube for example; they suggest that this generates a single unique powerful attractor. This deep attractor lends the viewer a stable conscious interpretation of the cube. Now consider an ambiguous figure like the Necker cube (fig 1.2 A) which has at least two representations a viewer can consciously interpret. Kornmeier and Bach would describe these two different interpretations as likely shallower attractors in close vicinity with relatively little barrier between them in perceptual space. Reversals might then signify an occurrence of when momentary instability of one attractor allows for perception to switch to the nearby alternative. This model allows for integration of bottom-up and top-down factors influences on reversals, both of which could cause instability of attractors, and offers a potential explanation for the observation of an RN in response to disambiguated stimulus variants, as these unambiguous stimuli could generate attractors that are close enough to one another in perceptual space via virtue of their visual similarity to be alternated between in a manner similar to their ambiguous counterparts.

Lastly, authors such as Pitts et al. (2007) and Kornmeier & Bach (2004, 2012), have raised the question whether the RN may be related in some way to the Selection Negativity (SN). The SN is an ERP component (identified by Anllo-Vento & Hillyard, 1996) which, like the RN, appears more negatively on trials with attended features (akin to reversal) than it does on trials without attended features (akin to stables). Perhaps the task relevance of reversal trials suggests them worthy of attracting attention. Were this the case, an SN would make sense. Perhaps then, the RN may be more about feature selection mechanisms than about switches in perception.

1.4.2 The LPC: Memory or Decision Making?

Another ERP component consistently observed in bistable reversal studies is the Late-Positive Component/Complex (LPC) which is a positive ERP component observed at central-posterior scalp regions generally ~350-650ms. The LPC (sometimes considered interchangeable or at least very similar to the P300) is found in response to both endogenous and exogenous reversals (Kornmeier and Bach 2006). Like the RN, this component is also a difference obtained when comparing averages of waveforms of reversal and stable trials, beginning around ~350ms after stimulus onset. The predominating general consensus on the LPC is that it reflects updating of the content of short-term working memory in the visual domain in an effort to report one's perceptual experience.

The LPC has been observed in studies where subjects report perceptual switches/stabilities as well as studies that require a response after each stimulus (Dien & Donchin, 2004; Davidson & Pitts, 2014) so the LPC is likely not solely related to reversals themselves. Furthermore, like the RN, the LPC is found in response to both endogenous and exogenous reversals (Kornmeier and Bach 2006).

Several studies have called into question whether the LPC *really* reflects context-updating, with Brezis et al. (2016) suggesting that while tied to the outcome of a memory decision, the LPC is not necessarily exclusively a marker of episodic recollection. It may instead reflect the strength of a signal that drives a decision as measured by expressed confidence. This suggests that the LPC is illustrative of a broader array of memory signals beyond context-updating in visual short-term memory. Yang et al. (2019) took this notion a step further suggesting that the LPC may be a marker of a participant simply deciding to respond or how to respond, which could change from trial to trial based on their perception and interpretation of previous trials.

1.5 Ambiguity within Other Paradigms and Other Modalities.

The Reversal Negativity and Late-positive component ERP signatures are not unique to the switching of bistable percepts but have also been observed in perceptual switching in binocular rivalry paradigms. Binocular rivalry is a perceptual paradigm where two different stimuli are presented simultaneously, one to each eye. In such paradigms the subjective perception spontaneously switches between the left eye and right eye stimuli (Britz and Pitts 2011).

Bistable perception also exists in other sensory modalities. Auditory stream segregation is a phenomenon in which a series of tones of two different pitches can be perceived either as two separate and simultaneous streams, or as one integrated stream alternating in pitch, with similar spontaneous switches between interpretations occurring as one listens to the stream (Snyder et al. 2015). An auditory analogue of the RN has been identified, however, in response to sequentially presented complex tones which can, in a bistable fashion, be perceived as ascending or descending in pitch, with discrete reversals happening at tone onset (Davidson and Pitts 2014). Perhaps the most pop culturally ubiquitous bistable paradigm is in the form of sound clips that go viral. Take for instance "Laurel and Yanny", a well-known bistable auditory stimulus that often recirculates online where one can hear either the name "Laurel" or the name "Yanny." In 2021, Alex Franklin, a student at Reed College, investigated the electrophysiological components of auditory bistability associated with the perception of the "Laurel and Yanny" sound clip as part of his senior thesis project. Franklin successfully found both an auditory RN and an LPC. This lends credence to the idea that the ERP signatures associated with bistable reversals may be conserved generally across sensory domains. Though not exactly a sensory domain in and of itself, a cognitive realm that is ripe with ambiguity, and which may lend itself to exploration using EEG is that of language processing. Therefore, the possibility of bistable linguistic stimuli and a corresponding Reversal Negativity in language is the focus of this study.

1.6 Lexical Ambiguity and The Challenges of Bistability

All known human languages are rife with ambiguity, and these ambiguities take many different and complicated forms (Youn et al., 2016). A primary form of lexical ambiguity is the homograph, a single word spelling that can have two or more non-overlapping possible meanings. Though it may seem counterproductive for there to be many potential meanings of a given word, in a sense language seeks out and relies on this kind of ambiguity because it makes communication more effective in that it gives people a way to convey a wide breadth of topics effectively and efficiently with a finite lexicon (Bartsch, 1984; Piantadosi, Tily, & Gibson, 2012; Ramiro, Srinivasan, Malt, & Xu, 2018; Schaff, 1964). This kind of lexical ambiguity at the single-word level rarely causes much trouble as humans tend to be quite capable of deriving the specific meaning from a homograph's context in a phrase or sentence. To an extent, a homograph is a stimulus that changes only in conception or interpretation, and not physically, and therefore, it is somewhat similar to a bistable figure.

Ambiguity at the sentence level can also resemble bistability. For example, a sentence like "The chicken is ready to eat" can conjure up ideas of a hungry chicken or a cooked chicken, with no physical changes to the sentence itself. Previous research on which the present study is based (see below) looks at linguistic analogues to bistable percepts at the sentence level.

1.7 Previous Attempt at Studying Linguistic Analogues to Bistable Percepts.

1.7.1 Ortego thesis

In 2018, Kevin Ortego, for his senior thesis, set out to investigate bistability in ambiguous language. Ortego investigated the reversal negativity in the linguistic domain through repeatedly presenting fully ambiguous sentences (such as "The chicken is ready to eat") which were then disambiguated by the presentation of line drawings depicting one of the two disambiguated interpretations of the sentence (i.e., a cooked chicken or a chicken eating food. See Fig. 1.4). On each trial, participants were instructed to report whether the drawing matched or mismatched their subjective interpretation of the previous ambiguous sentence. Because this ambiguous-then-unambiguous method of presentation had never been used in prior studies with bistable figures, Ortego also included a figure condition (e.g., the rat-man ambiguous figure, followed by an unambiguous man or rat) to confirm that this modified paradigm could effectively elicit the Visual Reversal Negativity (vRN) and visual LPC.

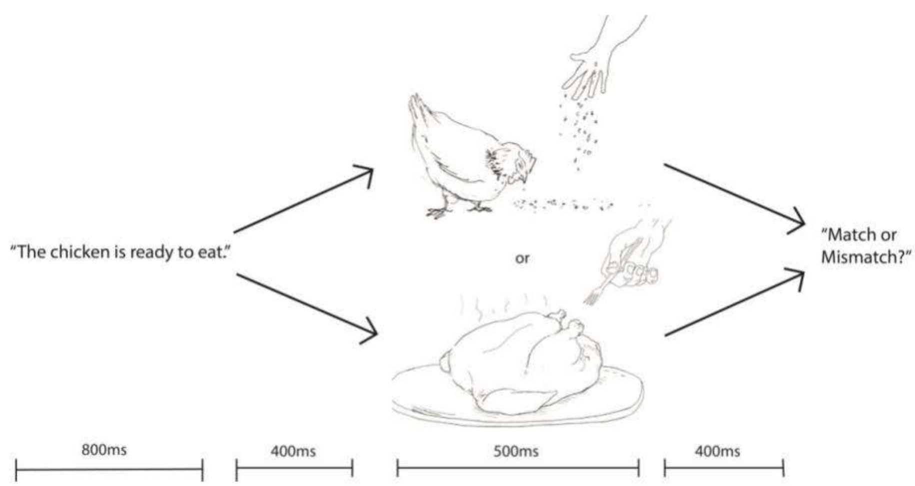

Figure 1.4 Example of Ortego's experimental design

Using this novel paradigm with ambiguous figures and comparing ERP recordings time-locked to the unambiguous stimulus in match vs. mismatch trials, Ortego first replicated the vRN (see figure 1.5; left side) and the visual LPC. Most importantly, when comparing match vs. mismatch trial now in the sentence condition, he also found two neural signatures of conceptual reversals similar to those of perceptual reversals: a large anterior negativity that occurs in the same time window as the visual RN (150-350ms) and a late positivity similar to the visual LPC that is observed in the 350-600ms interval (see figure 1.5; right side).

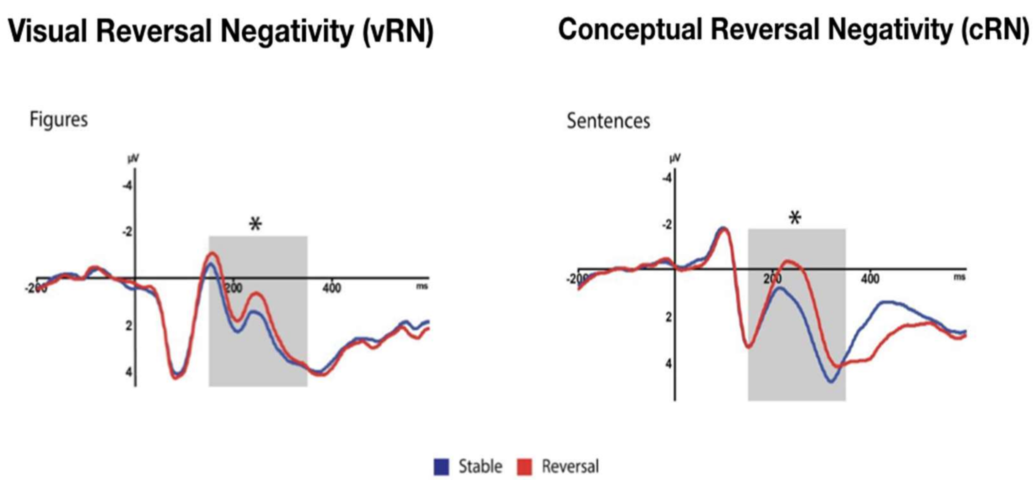

Figure 1.5 visual RN (in posterior areas: 26, 42, 43, 53, 54) and conceptual RN (in frontal areas: 3, 7, 9, 10, 18, 19, 21, 22, 33, 34) obtained in the figure and sentence conditions respectively (Ortego, 2018)

1.7.2 Possible Confound with Expectancy and Follow up Study.

In spite of these interesting results, Ortego's thesis left open the possibility that the "conceptual" RN (cRN) was not related to the reversal of interpretations per se but was instead an indication of the unexpectancy of seeing a drawing that is incongruent with the participant's mental representation of the ambiguous sentence. For instance, when reading the ambiguous sentence "The chicken is ready to eat", the participant might form the mental imagery of a cooked chicken and anticipate seeing the drawing of a roast bird, so the disambiguating drawing of a live chicken (which was the alternate option) may appear as a surprise and trigger the large anterior negativity observed in Ortego's study. Therefore, before concluding that there is a conceptual analog of the RN, we need to test the possibility that violation of expectation contributed to the reversal-like negativity in Ortego's study.

In 2020, Christy Lei, another student, attempted to follow up on Ortego's thesis with updated methodology and stimuli to test whether the cRN was in fact related to the resolution of ambiguity at a conceptual level or it was merely indicative of the violation of expectancy, which is presumably an entirely different cognitive process. Lei added trials that included unexpected stimuli to directly compare ERPs elicited by disambiguating stimuli in three types of trials: stable trials, reversal trials, and unexpected trials. Although Lei's design solves an issue with the initial study, and her findings could be better interpreted as supporting the proposal that the cRN may reflect a violation of expectancy, there are however other problematic issues with the use of images as disambiguating stimuli. Therefore, I designed the current study in such a way to eliminate both the use of images and any potential role of unexpectancy along the lines of what Lei found in the ambiguous-then-unambiguous approach to the reversal task.

1.8 Rationale and Hypothesis

As was the case in the two studies outlined above, the aim of the present experiment is to investigate whether the ERP signatures of reversals in bistable percepts and the underlying processing they reflect are domain-specific to the visual system, or whether analogous processes are involved as part of a more general perceptual phenomenon when confronted with ambiguity, in this case, in the domain of language. We will explore this question at the level of single, ambiguous words, as ambiguity at the sentence level is exponentially more complicated than ambiguity at the single-word level. With ambiguous sentences, one has to contend with multiple theories of heuristics and context-construction, not to mention the many different roles that verbs can play and the way they do so. In order to investigate the reversal negativity more directly in the linguistic domain, we will repeatedly present ambiguous words (e.g., Bank) and, along the lines of the more traditional reversal task, the participant will indicate on each trial, which one of two possible interpretations (e.g., "money" as in a financial institution or "river" as in the land alongside a river) they experience. We then will sort the EEG data from their responses into reversal and stable trials in order to compare them. Using a word-only approach will allow us to time-lock ERPs to the onset of the stimulus without the potential confound of switching between lexical and visual domains as in Ortego's thesis.

We hypothesize that if the paradigm of bistable switching uses similar neuroanatomical structures and pathways in the domain of language as it does in strictly visual percepts, then we will find similar ERP components, such as the RN and LPC, associated with the reversal trials. If the cRN found by Ortego is actually indicative of unexpectancy, we would not expect to find it within this paradigm. Of course, there is also the possibility that we may fail to find *any* significant differences, implying that the disambiguation of ambiguous lexical stimuli is not correlated with any sort of reversal signal.

1.8.1 Note on the N400

The N400 is an ERP component that occurs when we experience violations of semantic expectancy. The classic example being the sentence "I take my coffee with cream and *dog*" where the word "dog" elicits a larger negativity (around 400 ms) compared with the brain response elicited by an expected word like *sugar*. The N400 is also sensitive to semantic relatedness in a more general sense;

having been observed when presenting pairs of unrelated (e.g. flower-ham) versus related words (e.g. doctor-nurse) in isolation, and when presenting sentences word-by-word but with the critical unexpected word presented as an image (Nigam 1992; and Kutas & Federmeier 2012 for a broad overview of the N400 literature). As the N400 is a possible effect of lexical ambiguity research where any sort of contextual expectancy has the possibility of being violated, our experimental design will be structured in such a way to avoid confounds of unexpectancy.

Even though our study is attempting to eliminate the role of expectancy in its design, we may still see an N400. For example, perhaps a participant expects they will perceive the word "bank" as "river" but instead perceives it as "money." This sort of internal unexpectedness may show itself as some sort of N400. However, discerning such an effect is virtually impossible as it stands, because we would need a way to distinguish trials based on participant intent. Perhaps such an effect will be so strong that we cannot reliably produce a reversal negativity or perhaps the reversal negativity and the N400 are actually more indicative of the same underling mechanics behind response to unexpectedness. In fact, perhaps the LPC as seen in reversal trials in bistable paradigms could be thought of as a sort of backwards N400. This is to say ERP data in these kinds of studies is generally visualized as the mean amplitude data from stable trials being subtracted from the mean amplitude data of reversal trials. This leads to observing that late component as a positivity. However, if we instead subtracted the reversal trials from the stable trials, that identical brain process would be a negativity, one which may resemble the N400. As the N400 indicates violations in semantic representations, what this means exactly in terms of a reversal task would be that participants are perhaps, in some way, expecting to switch their interpretation of bistable stimuli. This line of thought is explored in more detail in the discussion section of this thesis, but especially given both the fact that we will be using word stimuli in our procedure *and* are attempting to eliminate unexpectancy from our procedure, the N400 is a crucial component to consider.

Chapter 2: Methods

2.1 Participants

A total of sixteen Reed college students with normal or corrected-to-normal vision and no history of brain injury or recent head trauma participated in this study. Due to complications with data collection that resulted in extreme electrical noise, two participants' data were excluded from statistical analysis. Participants were compensated with $20 for their participation. All procedures were approved by the Reed College Institutional review board.

2.2 Stimuli

Before we settled on the appropriate stimuli to use for the present experiment, we conducted a very informal pilot test to see if a reversal task was feasible in the domain of single, ambiguous words. The pilot test followed Ortego's above described ambiguous-then-unambiguous approach to the reversal task. The pilot test used twelve ambiguous words, each with two disambiguating meanings and one unexpected meaning (See Table 1). The notion of using unexpected meanings came as an effort to differentiate the difference between reversal and stable trials from any potential issues with unexpectancy from using Ortego's ambiguous-then-unambiguous stimuli presentation design. So, for example, the two disambiguating meanings of Bank were "money" and "river," and the unexpected meaning was "cloth". Word stimuli were taken from several different sources (Ishida, 2019; Onifer and Swinney 1981). Relative associative strengths (defined as forward strength: basically, a percentage of a polled population that responded to the given homograph with a specific definition) were then taken from the University of South Florida Association of

Free Norms (Nelson et al., 1998) and from the Edinburgh Associative Thesaurus (Kiss et al., 1973) databases. Not all associated word meanings were represented in both databases. Associated word meanings were arbitrarily categorized into "Meaning 1" and "Meaning 2" as dominant meanings of homographs varied across databases. All stimuli were presented in capital letters.

Table 1: List of proposed word stimuli
Included is each ambiguous words, their two associated meanings, and an unrelated meaning.

AMBIGUOUS WORD	MEANING 1	MEANING 2	UNRELATED MEANING
BANK	MONEY (*.799*, .026)	RIVER (*N/A*, .01)	CLOTH
BAR	PUB (*.014*, .01)	POLE (*.014*, N/A)	AUTUMN
CHANGE	DIFFERENT (*.165*, .041)	COIN (*.063*, .01)	SHED
COMPANY	BUSINESS (*.097*, .01)	FRIEND (*.097*, .03)	FEATURE
FIGURE	SHAPE (*.081*, .03)	NUMBERS (*.104*, .052)	MONTH
GLASS	EYE (*.010*, .02)	CUP (*.113* N/A)	TOWER
ISSUE	PROBLEM (*.062*, .01)	MAGAZINE (*.104*, .052)	COAST
JAM	STRAWBERRY (*N/A*, .146)	TRAFFIC (*.047*, .042)	SOURCE
KID	BOY (*.128*, .041)	GOAT (*.074*, .184)	SOUTH
PLANE	JET (*.014*, .01)	GEOMETRY (*.021*, .01)	FORCE
PURSE	SATCHEL (*.014*, N/A)	LIPS (*.014*, N/A)	ESTIMATE
SAGE	SPICE (*.014*, N/A)	WISE (*.095*, N/A)	THEME
(USF-FAN frequencies displayed in italics and EAT frequencies underlined. Unrelated meanings have no associative strength as they are unrelated to the ambiguous words at hand by design).			

This informal pilot study suggested that the switching task was indeed possible in the lexical modality, and the participants gave input that they thought a version of the task more akin to the traditional reversal task described in the introduction would be easier. The best option, in terms of investigating the RN, then became clear. In an effort to increase the present study's similarity to the standard procedure used in the bistable switching literature, we could adjust the general approach adapted from Kornmeier & Bach, 2004. The issue here, however, is that this version of the task is slightly more active and using 12 stimuli with hundreds of trials each, would likely yield diminishing returns as participants grew fatigued from the task, and as such we had to decrease the number of stimuli used. Using associative and forward strength data from the databases mentioned previously (see table 1) as well as input from the pilot subjects, three of the initial twelve ambiguous word stimuli ("bank", "company" and "issue") were selected for the final study (see Table 2).

Table 2: list of stimuli used in the present study.
Forward Strength based on the University of South Florida Association of Free Norms (italics) and from the Edinburgh Associative Thesaurus databases (underlined).

Ambiguous Word	Meaning 1	Meaning 2
BANK	MONEY (*.799*, .026)	RIVER (*N/A*, .01)
COMPANY	BUSINESS (*.097*, .01)	FRIEND (*.097*, .03)
ISSUE	PROBLEM (*.062*, .01)	MAGAZINE (*.104*, .052)

2.3 Procedure

All EEG recordings took place in an electrically shielded and sound-attenuated recording chamber with subjects seated ~75 cm from the screen. After participants were fitted with the electrode cap and gel, the experimenter guided participants through a familiarization block to ensure that they were aware of both possible interpretations of the ambiguous words and could practice switching between these interpretations. If a participant was unable to perceive both interpretations, the experimenter provided more defining information on the screen.

During the experiment, stimuli were flashed on the screen for 400ms each, followed by a 400-700ms interstimulus interval (ISI) prior to the onset of the subsequent stimulus. In each trial, participants were instructed to respond based on which interpretation of the ambiguous stimulus they activated. For example, in the case of the word "bank," participants would press one button if they thought of "money" and another if they thought of "river" (see Fig. 2.1). Participants were told to abstain from any response and wait for the next stimulus if the ambiguous word did not give rise to one clear interpretation. The three ambiguous words were presented in separate blocks, so each button corresponded to only one conceptualization at a given time.

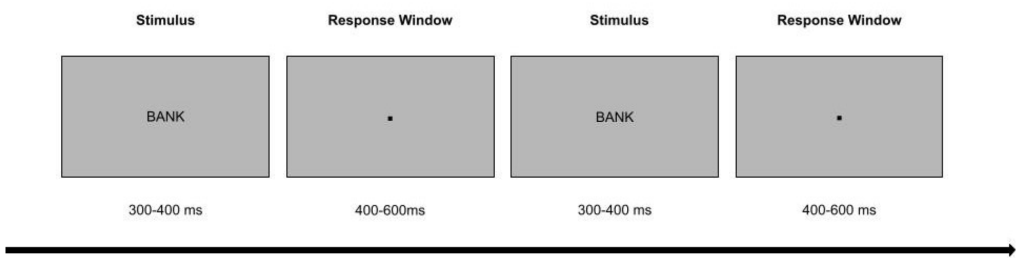

Figure 2.1 Schematic representation of one experimental Trial.

To explain the difference between a stable and a reversal trial, let us again consider the example of "BANK". A stable trial is defined as any time a response is the same as in the previous trial. In the case of "BANK" a stable trial would be a participant responding with the button corresponding to "money" when their previous response was also "money" (see figure 2.2 A). A reversal trial would consist of a participant responding with the button corresponding to "money" when their previous response was "river" (see Figure 2.2 B).

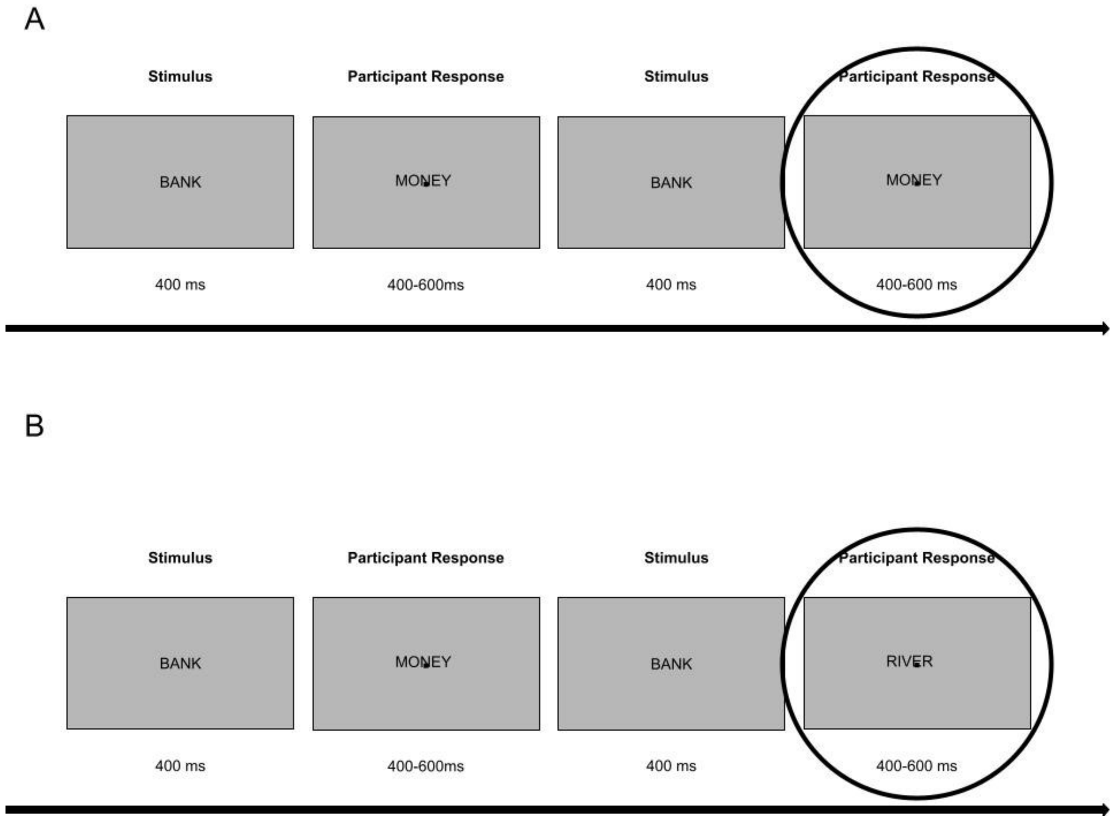

Figure 2.2 Schematic of trial types

(A) Example of a stable trial using the stimulus "BANK" The response defined as the stable trial is circled.

(B) Example of a reversal trial using the stimulus "BANK" The response defined as the reversal trial is circled.

The experiment consisted of three long blocks each consisting of stimulus presentations of one of the three stimuli. Each long block consisted of 300 trials, with breaks after every 50 presentations. At least one practice block of 40 trials was given at the beginning of each long block to acclimate participants to the task, the two new possible definitions, and the speed of stimulus presentation. Each recording session lasted 90-120 minutes, including setup time and cap/electrode preparation.

2.4 EEG Recording

For the EEG recording sessions, participants were fitted with a 64-channel electrode cap (Figure 2.3). An electrode (channel 64) placed on the face below the left eye (VEOG) was used to detect eye-blink artifacts, and two electrodes (channels 62 and 63) were placed adjacent to the left and right eyes (HEOG) were used to detect horizontal eye movements. Two additional electrodes (Channels 58 and "Ref") were placed on the left and right mastoids. Impedance levels at all electrodes were kept below 5kΩ. This was achieved with the use of a saline-based gel and gentle abrasion of the scalp with the wooden end of a Q-tip, in order to remove a thin layer of dead skin cells. Immediately after the session ended, usually within 3 hours of participants' arrival, caps were removed, and participants were able to wash their hair in the lab.

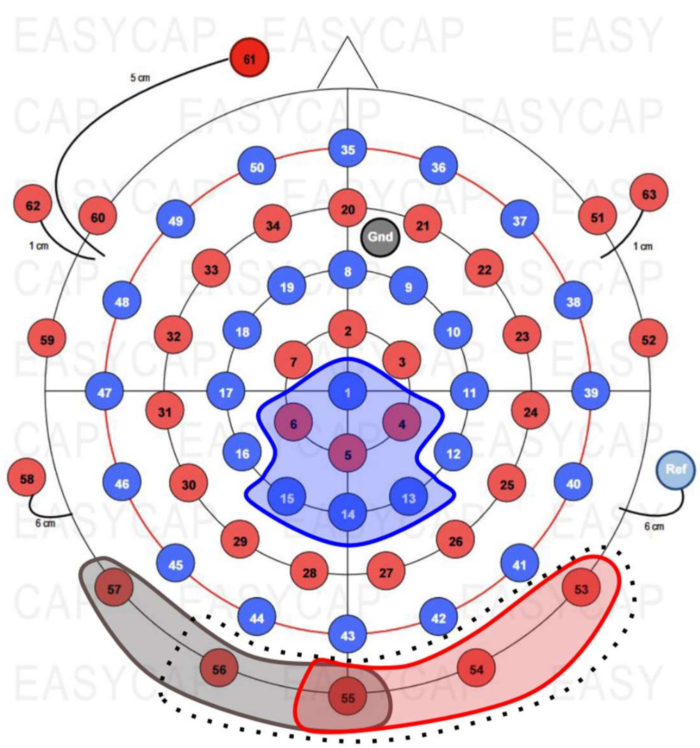

Figure 2.3 Scalp Electrode Atlas With Highlighted Pooling Areas.

The Area highlighted in gray represents the electrode pool investigated for the ERP component described in section 3.2.1, the area highlighted in red represents the electrode pool for the component discussed in section 3.2.2, the area highlighted in blue represents the electrode pool for the component described in section 3.2.3, and the area in the dotted region represents the component discussed in 4.4.1.

2.5 Data Analysis

All Electroencephalographic (EEG) data were processed using BrainVision Analyzer Software (Brain Products, Germany). Artifacts (blinks, eye movements, facial muscle noise, etc.) were rejected semi-automatically (on average ~20% of trials were rejected due to artifacts across all conditions EEG was recorded using a right mastoid electrode as a reference and re-referenced off-line to the average of the two mastoid electrodes. ERPs in each condition were time locked to the onset of the ambiguous word.

Chapter 3: Results

3.1 Behavioral Results

We observed an overall average reversal rate of 32.94% across all three stimuli and a similar average reversal rate for each of the three ambiguous words. A reversal rate of ~30% is standard for a reversal task, but it should be noted that certain participants experienced more extreme reversal rates (See Table 3). Based on the current data, we would expect that with many more trials and subjects, reversal rates for all three-word stimuli would settle in the 30-35% range. Participants were slightly money-biased when viewing the word "BANK" (money: 52.4%, river: 47.6%), slightly business-biased when viewing the word "COMPANY" (business: 52.1%, friend: 47.9%), and slightly magazine-biased when viewing the word "ISSUE" (magazine: 52.1%, problem 47.9%)

Table 3: Reversal rates for each individual stimulus and all stimuli together across all participants.

Subject	Bank	Company	Issue	All Stimuli
1	15.27%	20.17%	23.26%	19.05%
2	39.84%	24.48%	26.47%	29.98%
3	47.95%	46.48%	46.67%	47.09%
4	35.29%	75.00%	43.66%	44.20%
5	39.82%	33.33%	38.10%	38.07%
6	44.19%	36.43%	34.34%	38.59%
7	61.43%	60.87%	45.57%	56.02%
8	25.00%	9.52%	16.00%	17.36%
9	32.86%	21.84%	18.56%	23.62%
10	25.22%	9.09%	10.91%	15.03%
11	46.05%	33.33%	29.47%	36.23%
12	26.32%	48.33%	40.91%	33.80%
13	33.85%	42.48%	50.00%	42.26%
14	19.05%	22.83%	16.88%	19.83%
Average	33.46%	34.58%	31.49%	32.94%

3.2 Electrophysiological Results

In order to evaluate the collected data in a manner that would allow us to answer the questions we had about linguistic ambiguity, for each participant, their EEG data was segmented into "Stable" and "Reversal" trials following the procedure indicated above in section 2.3.

As this single bistable word paradigm is a novel approach to finding neural correlates for lexical ambiguity processing, we cast a wide net looking for ERP signatures. We kept a keen eye toward Ortego's previous findings of the "conceptual" reversal negativity and LPC, but also looked for previously solidified components such as the vRN. As we were not certain to expect a replication of Ortego's findings, a replication of the findings along the lines of a more traditional reversal task with bistable figures, or something else entirely, regions of interest (ROIs) were determined by investigating our difference maps. These maps are created by subtracting stable from reversal ERPs to show a distribution of differences on the surface of the scalp (see below).

Based on our hypotheses, on previous studies, and a visual analysis of ERP difference maps, we measured the mean amplitude of individual averages for each condition in two-time windows. We were primarily interested in the times and ROIs corresponding to the RN and LPC as mentioned in the introduction.

We then visually inspected the EEG data for the presence of language-related ERP components such as the N400 and P600, or any other effects. Based on the examination of scalp maps of all three components, we defined regions of interest on the scalp in which to compute mean amplitudes of the different ERP signatures during specified time windows (see the corresponding three regions of interest or ROI on Fig. 2.3).

Below I report our findings in turn for each of 3 different observed ERP components.

3.2.1 Early Posterior Negativity

The grand average (across all participants) of waveforms elicited by stable trials is shown below in Fig. 3.1 A (shown in black), overlapped with those elicited by reversal trials (shown in red) in the left-occipital areas of the scalp.

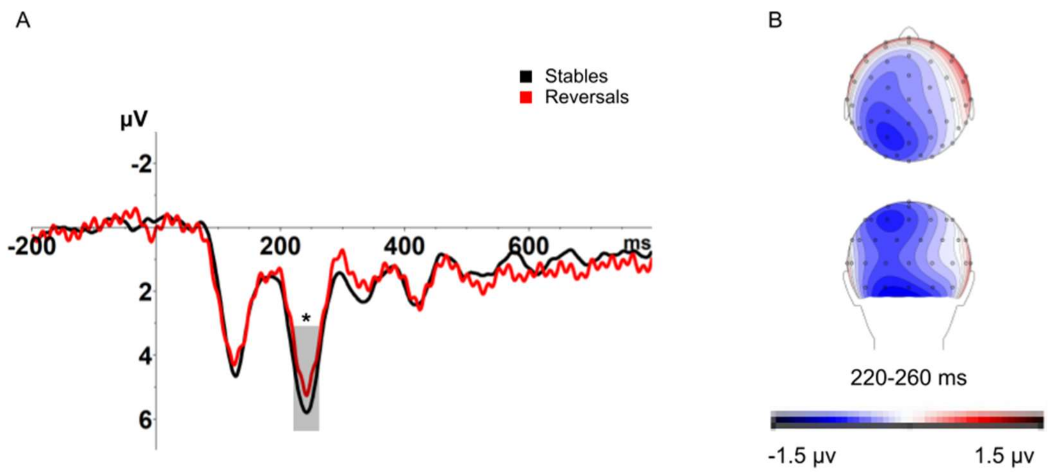

Figure 3.1. Early Posterior Negativity and Potential Dipole ERPs and Difference Maps.
(A) Average ERPs elicited by stable and reversal trials in a pool of posterior electrodes (55, 56, 57). The highlighted area shows the 220-260 ms time window selected to compare the difference in mean amplitude between stable and reversal trials. In this and future figures, the star denotes a statistically significant difference between conditions in that time window, $p < 0.05$.
(B) Difference maps showing the scalp distribution of the difference between stable and reversal trials from 220-260ms.

The map view of the scalp (See Fig. 3.1B) suggested a left-occipital negativity associated with reversals from around 220ms-260ms. To investigate this effect, mean amplitudes for this time window were computed from a pool of three electrodes (55, 56, 57) corresponding to the left-occipital ROI. A one-tailed t-test confirmed that the mean amplitudes elicited by reversal trials (M=4.53µV SD=2.25) was significantly more negative than that elicited by stable trials (M=5.22µV SD=2.43), $t(13) = 1.85$, $p < .05$.

3.2.2 Second Negative Component

Fig. 3.2A below shows the overlapped grand average of ERPs elicited in stable trials (shown in black) vs. those elicited in reversal trials (shown in red) in the right-occipital areas of the scalp.

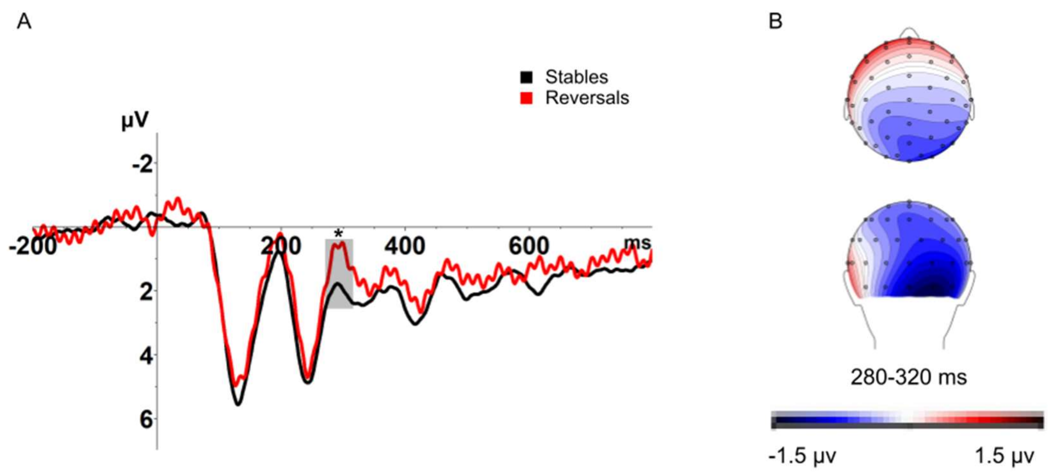

Figure 3.2. Later Right-shifted Posterior Negativity and Potential Dipole ERPs and Difference Maps.

(A) Average ERPs elicited by stable and reversal trials in a pool of posterior electrodes (53, 54, 55) The highlighted area shows the 280-320 ms time window selected to compare the difference in mean amplitude between stable and reversal trials. Stars denote a statistically significant difference in that time window, $p < 0.05$.

(B) Difference maps showing mean amplitude difference between stable and reversal trials averaged across a representative time window from 280-320 ms.

The map view of the scalp (See Fig. 3.2B) suggested a right-occipital negativity from around 280ms-320ms. To investigate this effect, mean amplitudes for this time window were computed from a pool of three electrodes (53, 54, 55) corresponding to the right-occipital ROI. A one-tailed t-test confirmed that the mean amplitude elicited by reversal trials (M=0.83μV SD=1.84) was significantly more negative than that elicited by stable trials (M=2.00μV SD=2.28), $t(13) = 1.96, p < .05$.

3.2.3 LPC/P3

Fig. 3.3 below shows the overlapped grand average of ERPs elicited in stable trials (shown in black) vs. those elicited in reversal trials (shown in red) in the centro parietal areas of the scalp.

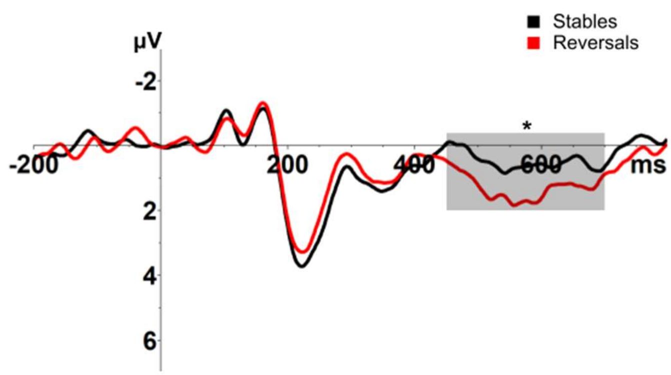

Figure 3.3. LPC/P3 ERPs

Event-Related Potentials obtained from a pool of centro-parietal electrodes (1, 4, 5, 6, 13, 14, 15) ROI corresponding to the P3. Stars denote a statistically significant difference in that time window, $p < 0.05$.

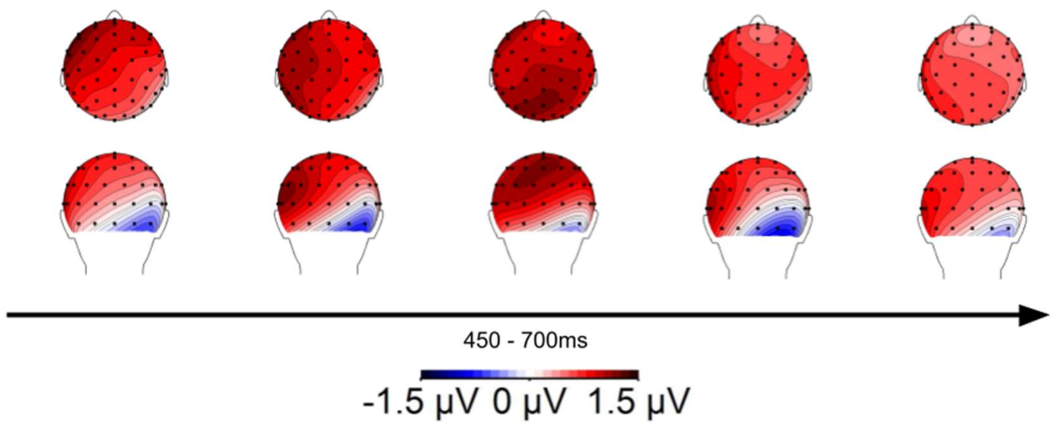

Figure 3.4 LPC/P3 Difference Maps
Difference maps showing mean amplitude differences between stable and reversal trials during peak effect.

The map view of the scalp (See Fig. 3.4) suggested a centro-parietal negativity from around 450ms to 700ms with the strongest effect around 560-600 ms. To investigate this effect, mean amplitudes for the time window between 450-700 ms were computed from a pool of seven electrodes (1, 4, 5, 6, 13, 14, 15) corresponding the centro parietal ROI. A One tailed dependent means t-test confirmed that mean amplitudes in reversal trials (M=1.36µV SD=1.87) were significantly more positive than those in stable trials (M=0.49µV SD=2.31) in the given ROI, $t(13) = -2.10$, $p < 0.05$. The timing, location, and polarity of this ERP are consistent with the LPC or potentially a late P300 (P3).

Chapter 4: Discussion

4.1 Summary of Results

In summation, we failed to find Ortego's cRN. However, two occipital negativities were found in reversal trials compared to stable trials. Whether these components are indicative of a single RN, or an RN and an additional negativity is discussed below. In addition, a large centro-parietal positivity was observed in reversal trials, compared to stable trials between 450-700 ms. We identify this ERP component as the LPC. I now discuss each of these findings in detail.

4.2 Results in Context of Ortego's thesis

We failed to replicate the finding of the cRN as described in Ortego's previous research. However, we were not exactly *expecting* to find this ERP signature, and in fact, these results may confirm that Ortego's findings were more indicative of unexpectancy. As such, the cRN reported by Ortego may be either a confound of unexpectancy or some other signifier of ambiguity processing at the sentence level versus the word level. However, like Ortego, we observed an LPC, although in the present study the LPC was observed roughly 100 ms later than in Ortego's study, between 450-700 ms.

The results observed in the present study appear to fall more in line with the extant literature with bistable, visual paradigms. On first impression, this result was rather shocking because this task "feels viscerally" different from such visual reversal tasks. The ERPs associated with those tasks, like the RN and LPC, are thought to be illustrative of neural processes relating to changes in *visual perception*. These visual stimuli *appear to change*. In contrast, the task in the present study may be better described as aimed to investigate similar changes in *interpretation*. In particular, the word stimuli do *not* appear to change, they sit stagnant, constantly flashing on the participant's screen with no change on *physical appearance*, objective nor subjective.

As briefly noted in the introduction, Bachmann and Aru (2023) recently described the contrast between different interpretations of conscious contents of the *already consciously experienced* stimulation as "conscious interpretation." The model of the present study has pushed the notion of conscious interpretation almost to its breaking point. These stimuli, in contrast to other paradigms with bistable figures, could be considered as changing *even less* in the domain of vision.

In regard to this previous research, the question then becomes whether we are truly observing activity suggesting that the two interpretations of the ambiguous words are being processed and alternated between, in the same way as the perception of bistable figures. There is then an even broader question: if the observed ERP is a reversal negativity, is our current understanding of the reversal negativity as a marker of changes in visual perception flawed? Could the RN be indicative of more broad processes of ambiguity processing?

4.3 Is the Early Component a true RN?

Is the component associated with reversals observed in a pool of three left-occipital electrodes between 220-260ms an RN? The time window, polarity, and topography all fall into place when compared with a standard RN, though this reversal appears slightly more concentrated to the left side of the scalp. The factors in favor of this component being a reversal negativity are as follows: (1) it occurs around the same time as previously observed RNs, (2) it is a primarily occipital effect. However, in contrast, there are some potential issues, i.e., this component is left-occipital whereas in reversal tasks with figures (including Ortego's thesis) the RN is often shifted to, or larger in the right hemisphere.

If our RN has this left-shifted occipital component, why could that be? It may be enticing to assume that this left-skew has something to do with language processing. This line of thought could come from the simplistic approach to understanding the brain's processing of language, which suggests that cognitive mechanisms behind the processing of language are generally thought to be anchored in the left hemisphere of the brain. This, however, is a generally outdated framework, and the truth is considerably more complicated than "Language happens on the left side of the brain." There are many language processing structures that exist in both hemispheres of the brain, such as the superior temporal sulci and gyri which are involved in phonology processing in spoken word recognition. One of the leading schools of thought is that language is processed and organized into two streams that exist within different regions of the brain, the ventral (lower) and dorsal (or upper) streams. This framework is often described as a "What/how" structure where language processing goes through the ventral stream (the "what") where semantic representations, such as word meaning, are processed, and the dorsal stream (the "how") where more detailed processes occur in many factors of language processing, like deciphering sound and planning articulation. However, this dorsal/ventral stream notion does not explain our left-skewed occipital finding as many of the regions in these streams exist bilaterally in the brain. There are certain regions in the ventral stream, like the anterior temporal lobe, which are only extant in the left hemisphere, but these regions seemingly are more entrenched in the integration of semantic knowledge. As such, these regions seem more involved in the processing of more complex lexical ambiguity at the sentence level.

There is another model of visual word recognition that may explain the observed left-shift in our EEG readings. Some researchers argue that the visual word form area, a region located within the rather large left occipitotemporal gyrus (sometimes called the left fusiform gyrus), is the brain's "letterbox," processing written letters and words and then transmitting this lexical information to higher-order language regions for linguistic processing (Dehaene & Cohen 2011; Cohen et al., 2002; Glezer et al., 2009; Plaut & Behrmann 2011). The visual word form area responds to learned letters over other meaningless letter-like shapes (Price et al., 1996b), it responds to both upper and lowercase letters even when visually disparate (Dehaene et al., 2001), it is accessed automatically as even when participants are unaware of words being presented, the area is activated (Dehaene et al., 2001), and finally electrophysiological data suggests that this area is activated early; around 150-200 ms (Bentin et al,. 1999). There is some doubt as to whether the visual word form area is indeed indicative of a stored lexicon of known letters and words, and more recent studies suggest that the visual word form area function is characterized by distinct circuits for integrating language and other cognitive functions like attention (Chen et al., 2019). Taken all together the theory of the visual word form area is an enticing potential explanation for the observed left-skewed occipital negativity. Perhaps some of the circuitry underlying interpretive switching in bistable word stimuli must occur in the visual word form area. This would explain why this region is generally not significant in the traditional RN observed in visual bistable paradigms.

Perhaps, too, this component is something akin to an SN. Participants could have had heightened attention on reversal trials. Why? One possible explanation, which is a bit of a recurring theme in this discussion, is that despite the instructions given, participants could not help but predetermine how they would perceive the word. That is to say participants, perhaps even blind to their own thought processes, already determined their interpretation before the stimulus presentation. If that is the case, it may have taken more attending to when they *chose* to switch. That attentional differential might manifest as the SN.

Lastly, there is a fairly obvious difference in this reversal task compared to those in purely visual domains. Look back at one of the bistable figures shown in the introduction, when your perception of the image shifts, you will likely experience some sort of visceral reaction. It almost feels shocking or like some kind of trick, or at least some surprise. Now look at the word "Bank" and switch between the definitions of "money" and "river". I strongly doubt you experienced any feeling of visceral shock. Perhaps this dichotomy of experience is reflective of two completely and obviously different patterns of neural activity. However, one could argue that this differential in feeling is mostly due to differences in novelty. We encounter homographs every single day and constantly use context clues to ascertain its appropriate meaning. This is very different from the experience of seeing the Necker cube. Perhaps if we interacted with bistable figures as commonly as we interact with homographs, their perceptive flexibility would not elicit such a visceral response.

Nonetheless, we believe that this observed negativity is an RN in the domain of language. This could mean that these stimuli truly behave as bistable phenomena. Although it might be more accurate to think of them as multisable phenomena because the words used in the present study in actuality have more than two possible meanings. However not to get ahead of ourselves, this idea really is just acknowledging the concept of polysemy, which is the notion that any given word can hold many meanings. This is just another way in which the stimuli in the present study are inherently different from the bistable visual stimuli seen in previous studies. The question then for future research (if this observed RN is indeed replicable), would be whether the RN is truly indicative of a change in perception, as it is currently defined. This must be confirmed because the conventional understanding of the RN does not account for stimuli that always “look” the same. Of course, a bistable figure stays constant, but the way it *looks* changes. In contrast, a word always looks the same. Alternatively, perhaps there is some kind of shared role of mental visualization (imagery) involved in the present study and in figure-based reversal tasks.

4.4 What is the Second Negativity?

Across all stimuli a right-occipital negativity was observed between 280-320 ms for reversal trials in the right-occipital regions of the scalp. The earlier RN, the concurrent/subsequent parietal negativity and this component all occur within 160ms and may be part of a broader reversal process. Allow yourself to be taken on a bit of a hand-waving journey. What if the initial left-skewed negativity signifies language processing? Though one of the limitations of EEG research is that it can be difficult to ascertain where in the brain a given signal is actually originating. Given that the effect of the RN here is so concentrated to the left-occipital regions, it may be suggestive of language processing as described above. So, following that logic, perhaps from ~220-260ms post stimulus onset during a reversal trial something happens in the visual word form area. Then perhaps the following right-shifted occipital negativity is indicative of some kind of process of visualization. Participants were informally asked how they approached the task, and some participants did indicate that they visualized the word meanings before responding. However, there was no standardized or collected data in this regard, so this will have to be explored in future research where investigators actively report participation strategies. If what is being proposed here is in fact true, then investigators could expect to find this right-shifted negativity in participants who report visualization in their approach to this lexical reversal task.

Alternatively, maybe the right shifted occipital Negativity is actually the true analogue to the reversal negativity, and for a reversal to happen in the domain of language, the stimuli must first go through some sort of identification processing that takes place in clusters of neurons located occipitally and parietally on the left side of the scalp. This of course is highly speculative. To our knowledge no previous reversal task studies have discovered a similar "moving" RN like this.

There are of course alternative explanations considering that Ortego and Lei's studies found nothing like this in reversal trials and it does not appear to be a common finding in the extant consciousness literature dealing with reversal tasks. What then can we make of such a signal? If it is a previously identified ERP, it could be a N270 which has been identified as being involved in endogenous conflict processing in the brain (Wang et al. 2000, Cui et al. 2000). The present study is deeply rooted in top-down effects. That is to say a participant's understanding of the stimulus affects their interpretation of that stimulus. However, the N270 is a marker of bottom-up effects (i.e., stimulus-driven) so it is unlikely to be what we are actually observing. Furthermore, the N270 is primarily elicited by a mismatch in a Same-Different Judgment (SDJ) task. In these tasks, a first stimulus (S1) is presented, followed by a second stimulus (S2) a few hundred milliseconds later that is either the same (match) or different (mismatch) in some given dimension. Our stimuli, of course, stay exactly the same with each presentation. Thus, unless the change in perception during a reversal trial in the present study is *so strong* that there is some sort of mismatch causing some sort of conflict, it would seem rather unlikely to be an N270. However, neither of these potential explanations are verifiable without following up on this study.

4.4.1 One Component or Two

Another possibility is that this negativity and the early negativity are actually one long ERP component that is shifting around slightly. Figure 4.1 below shows that if one includes all but the leftmost occipital electrodes (53, 54, 55, 56) the reversal trials from 220-320 ms in total are significantly more negative ($M=2.11\mu V$ $SD=1.41$) than the stable trials ($M=3.17\mu V$ $SD=2.08s$ $t(13) = 2.35$, $p < .05$). However, if the leftmost occipital electrode (channel 57) is included, the fact that it becomes less negative across those 80 ms greatly reduces statistical significance.

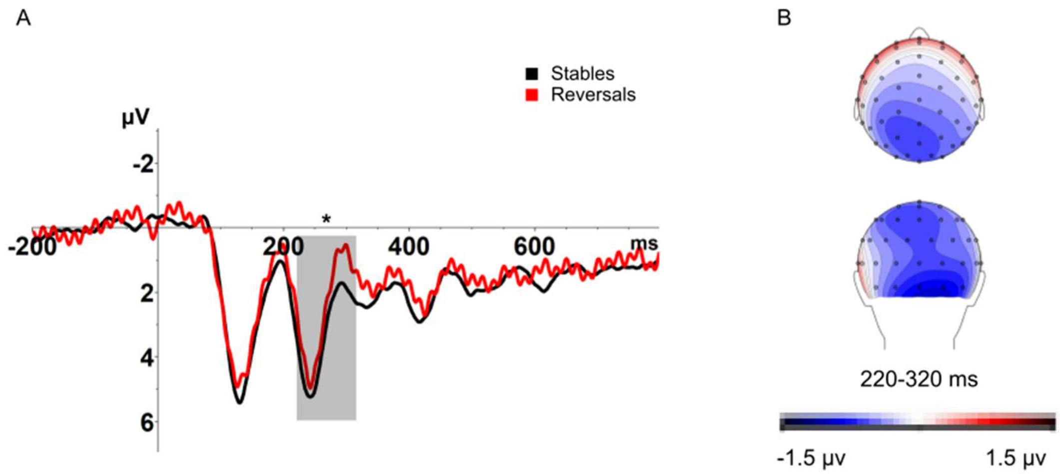

Figure 4.1. Proposed RN ERPs and Difference Maps.
(A) Average ERPs elicited by stable and reversal trials in a pool of posterior electrodes (53, 54, 55, 56). The highlighted area shows the 220-320 ms time window selected to compare the difference in mean amplitude between stable and reversal trials. Stars denote a statistically significant difference in that time window, $p < 0.05$.
(B) Difference maps showing mean amplitude difference between stable and reversal trials averaged across a representative time window from 220-320 ms.

The tidiest explanation would be that these two early components are in fact *one* RN. However, when you look at the scalp maps of this component in twenty-five second increments (Figure 4.2) this component appears to move ever so slightly left-to-right topographically. Perhaps the RN in this context shifts from left to right after lexical identification has occurred in various occipital-left regions, but as far as we can tell, a moving ERP component like this is not a finding widely reported in the extant RN literature. Future research, taking visualization strategies into account, will have to determine whether this is a single, shifting RN or if one of these negativities is an RN and the other is something else like the alternative explanations given previously in discussion section 4.4.

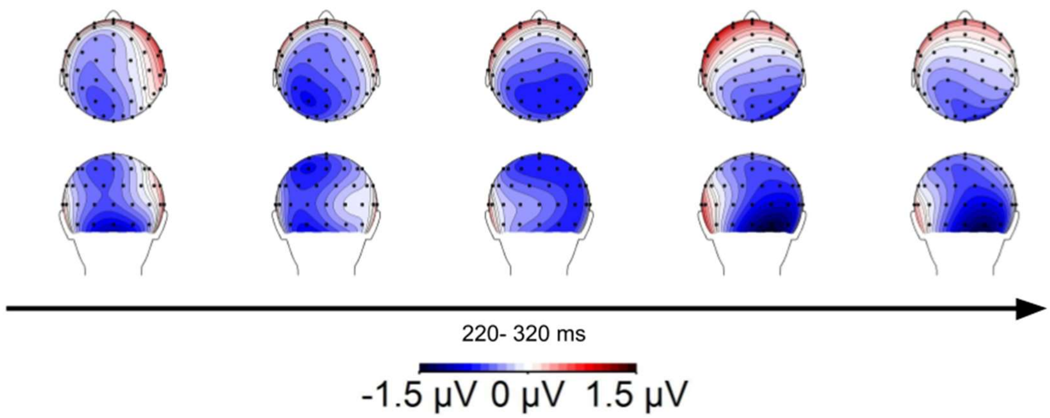

Figure 4.2 Difference maps corresponding to the observed early negativities.

Difference maps showing mean amplitude difference between stable and reversal trials averaged across a representative time window from 220-320 ms incrementally represented in 20 ms segments.

4.5 A Note on the LPC/P3

The question then arises of whether the late positivity observed across all reversal trials was in fact the LPC/P3 or some other ERP component. The latency of this component was very similar to Ortego's reported LPC and occurred in very similar, parietal and central regions of the scalp, however the latency was shifted about 100 ms later but still within the timeframe for the LPC or a later P3. The differentiation between the LPC and the P3 remains a subject of debate, with some authors using the terms interchangeably or even using alternative terms such as the P530 or P600. Therefore, for the purpose of simplicity and readability, in the present study, we will refer to the late positivity observed as the LPC. Regardless, we are comfortable concluding that this component, by far the most significant finding in the present study, is indeed the LPC.

But then what does it mean in the context of our study? The classic view of the LPC is that it represents the processing of episodic details related to the previous stimulus encounter. This recollection is often framed as a sort of context updating. That would make sense here because we identified the LPC during reversal trials. Following this interpretation, when someone's interpretation of an ambiguous word stimulus changes, there is a new context to the stimulus that needs to be processed in the brain. However, as previously noted, there have recently been attempts to recontextualize the LPC and what it exactly represents cognitively.

Yang et al 2019. suggest that the LPC may be a marker of a participant simply deciding to respond or how to respond, which could change from trial to trial based on their perception and interpretation of previous trials. They employed judgements of cumulative lifetime exposure to object concepts, and judgements of cumulative item recent exposure (which they refer to as frequency judgements) in a study-test paradigm. A comparison of ERP signatures in relation to degree of prior exposure across the two memory tasks and the study phase revealed that LPC elicitation tracked cumulative exposure regardless of whether accumulation of memories happened recently in the lab setting or over the participants' lifetimes. Most importantly they identified that this effect was only present when the memory judgment at hand required decision making in some relevant dimension. This explanation also neatly fits into our paradigm. Though in the instructions for the present study, participants were asked not to predetermine their interpretation of the ambiguous stimuli (i.e., they were told not to switch in a predictable sequence like money-money-river-river etc.) there is still undoubtedly a sort of decision-making process that happens in this or any reversal task. It is understood that there are two ways to perceive any of the given stimuli so even if a participant is keeping their mind as open as possible there may still be an element of choice. In other words, participants may be deciding, whether they are aware or not, which definition they interpret on each trial. If the LPC is truly an indicator of decision making, it would then come as little surprise that we observed an LPC in the present study.

Lastly, as a quick aside it should be noted that this component is considered a positivity because we are subtracting stable from reversals trials, as is common practice in this sort of study. However, the story changes when one considers this component the other way around. What if instead we think of it as a negativity elicited in stable trials? In such a case this strong centro-partial negativity seen at 450-700 ms would more closely resemble the well-established N400. The N400 of course is a marker of semantic unexpectancy seen across the board in language paradigms. But what would be unexpected about a stable trial? How could a stable trial be unexpected when nothing changes? Perhaps participants *expect their interpretation to switch*. Though participants were instructed to keep their mind open to both interpretations on every trial, and respond indicating which interpretation they experienced, perhaps being a participant in one of these studies suggests to them that the name of the game is switching. So perhaps when interpretation remains constant, there is a level of unexpectancy there manifesting itself as an N400, or N400-like component, that is routinely overlooked as we tend to consider this component as being elicited by reversals in perception.

And again, just to drive this point into the ground, the N400 is generally believed to reflect a violation in semantic expectancy. Such violations in the present study should not occur unless somehow the design leads to internal semantic incongruities for participants, but no participants expressed the detection of any sort of semantic violation. Of course, perhaps they *did* experience a semantic violation but just were attentionally unaware of such an effect.

All in all, our LPC is as solid as any LPC in the extant literature; that is to say it is definitely *there* and means something likely about either recollection/context updating or some kind of decision-making process, but we are yet to understand what *exactly* it represents and where in the brain it originates. Our own RN finding forces us to think about it more broadly. In contrast, observing an LPC in the context of our study makes sense regardless of which interpretation of the LPC you hold dear.

4.6 Role of Meaning and Participant Report

Something to consider when assessing between-participant variation is that participants may have employed different strategies while performing this task. The instructions asked participants to stay open to both interpretations of the word and try to perceive one of those two interpretations on each trial. Instructions advised against forming a predetermined sequence of switching. How participants carried out this task could have varied wildly. Some participants noted that they were visualizing a mental picture corresponding to one of the two meanings on each trial. Others mentioned that they were more passive and let the conceptualization of each stimulus wash over them. No data was formally collected to assess strategies, and as such, a systematic investigation of the effects of task strategy on individual results is currently non-viable. Particularly noteworthy is the potential role of mental imagery.

Individuals' ability to produce mental imagery varies, and questionnaires like the Vividness of Visual Imagery Questionnaire (VVIQ) attempt to evaluate this subjective phenomenon (Marks, 1973). Functional magnetic resonance imaging (fMRI) studies comparing mental imagery to actual perception have revealed considerable overlap in activation in the frontal, temporal, and anterior parietal regions, with variations between the two tasks mainly found in posterior parietal and occipital regions (Ganis et al., 2004). Furthermore, visual cortex activity has been found to be related to individuals' imagery vividness, as measured by the VVIQ (Cui et al., 2006). Recent research suggests that people with low imagery vividness activate a more extensive network of brain regions than those with high vividness, whose activations are more concentrated. Many of the areas activated only in the low vividness group show a negative relationship with vividness of imagery (Fulford et al., 2017).

Event-related potentials have limitations for studying mental imagery due to the challenges of precisely timing the fully internalized and temporally extended process of generating such an image. Nevertheless, Farah et al. (1989) recorded EEG data time-locked to the onset of visually presented words in two conditions, one where participants silently read the word and another where they created a mental image of the word's referent. Compared to the passive reading condition, the imagery condition elicited a late, slow positivity from ~600-1000ms, which was strongest at occipital and posterior temporal electrode sites and was lateralized to the left hemisphere. Shen et al. (2015) found a frontal negativity associated with imagery between 200-750ms using word-by-word presentations of literal versus abstract sentences. The amplitude of this effect was correlated with participants' VVIQ scores, with high-VVIQ participants showing a similar imagery effect when reading sentences with unfamiliar or familiar metaphors compared to literal sentences, and the effect being more prolonged for unfamiliar metaphors. However, low-VVIQ participants did not display any imagery effect in either metaphor condition, although they showed a posterior N400 effect for unfamiliar metaphors. The authors suggest that the frontal effect observed in high-VVIQ participants could reflect recruitment of frontal sensory-motor areas during visualization. However, caution should be exercised in interpreting this effect because an ERP effect's scalp location does not necessarily imply that the neural generators responsible for the effect are located in a similar brain region (Luck, 2014).

This previous research suggests that ERP signatures are associated with mental imagery, although they are tenuous. Additionally due to the novelty of our approach and the lack of previous studies employing a bistable approach to lexical ambiguity it is difficult to determine the role that mental imagery played in the current experiment. It is plausible if not likely that differences in mental imagery could have influenced some of the observed effects. Future research should explore imagery and general task strategies. Perhaps the changing topography and latency of the negativities observed was due to a sample containing participants who employed several different task strategies but experienced similar general reversal processes, nonetheless. To gain a fuller understanding of the observed ERP components a future study could split up participants into those who claim to visualize the different meanings of the ambiguous words and those who do not. Several hypotheses could suggest that increased vividness in mental imagery strengthens, weakens, or has no impact on the behavior being investigated in the present purely lexical reversal task, so the only way to know for sure is through more experimentation.

4.7 A Note on Interstimulus Comparison

One aspect of data analysis that has been absent from the present study is a comparison of the results between the three different word stimuli. The reason for this is that there are simply not enough number of trials associated with each stimulus to obtain a clear ERP signal. The design of the present study was rather tedious for the participants, and we found that it was difficult to keep participants engaged beyond 900 total stimulus presentations, 300 per each stimulus. Of those 900 total stimulus presentations participants averaged 82 reversal trials to 172 stable trials. However, for each stimulus trial totals are obviously much lower with 30, 26, and 25 average reversal trials for "BANK", "COMPANY" and "ISSUE" respectively, compared to 58, 57, and 56 stable trials. With only fourteen total participants and so few trials per stimulus block, mean amplitude data from the EEG recordings of reversal trials are much more prone to noise and extreme results that would likely settle over time as more data from more participants is collected. One could try to counteract this by comparing a smaller sample of stable trials in each stimulus, for instance only comparing 30 stable "BANK" trials to the 30 reversal trials to give the reversal trials more weight (or more accurately give the stable trials less weight). However again, the sample size is just so small that we would be uncomfortable drawing any major conclusions from this data. However, we can report that the general ERP data follows very similar trends across stable and reversal trials for the three different stimuli. But again, this should be taken with a grain of salt due to sample size.

Though we do not believe this necessarily invalidates our findings combining stable and reversal trials across stimuli, future research must investigate whether different kinds of ambiguous word stimuli yield different results. For instance, all the homographs used in this study have possible definitions that are verbs. Though we focused on noun definitions because ambiguity at the verb level adds layers of complexity, future research may investigate whether there is any difference in the effect with ambiguous verbs versus nouns. This can be investigated in a number of ways. A hypothetical study could be conducted to compare bistability across parts of speech in words such as "BANK." As in the present study, one group of participants would be asked to interpret between two noun definitions, such as "money" and "river". Another group of participants could be asked to interpret between one noun definition and one verb definition, such as "Money" and "heap" (as in the action of piling objects into a mass). Finally, a third group of participants could be asked to interpret between two verb definitions, such as "heap" and "tilt" (as in the action of an airplane tilting or turning). Based on the present study we do not have any particular predictions as to what the results of such a hypothetical experiment may be, *but* it would not be surprising if the noun-verb switches are different in some way than the noun-noun or verb-verb switches. Though the noun-verb condition would still be a bistable percept of sorts as it is representative of two simultaneous representations being held mentally at once, it feels almost beyond the confounds of bistability as it is normally experimentally probed. Additionally, when verbs get involved, they can trigger morphosyntactic processing, which is related to the handling of abstract verb-based grammatical information. As such, a study looking at bistability across these different parts of speech might well see significant differences in the neuronal structures involved when participants interpret an ambiguous word as a verb as opposed to a noun. Of course, this is all conjecture, and it is an interesting avenue of future research both in terms of helping us understand lexical ambiguity but also the limits of bistability processing more broadly.

4.8 Limitations

The main glaring limitation of the present study is the small sample size of only fourteen participants. As mentioned above, the small sample size combined with the fact that—due to the relative difficulty of this task—our total stimulus presentations were on the low end of the norm for this kind of reversal task, there was nowhere near enough data to identify statistically significant effects at the level of each of the three stimuli. Additionally, this small sample size questions the validity of certain behavioral comparisons, such as participant bias toward certain definitions. As such, data across stimuli was processed together for all the analysis in this thesis, and though the reversal rates and biases were similar across stimuli, all behavioral and EEG data must, unfortunately, be taken with a grain of salt.

Though we truly believe the EEG data averaged across all stimuli is indicative of the reversal mechanisms at hand, 14 participants is low for the confidence needed to identify statistically significant effects. EEG studies like this really need 25-40 participants to be taken seriously. The small sample size may contribute to the confusion as to whether the early negativities observed were truly one effect or two. With more data points that distinction may well have become clear. Ultimately, since the observed data is rather compelling, and as some preliminary analysis of weighted averages suggests that the differences observed might not just be the result of a small sample size, future research should replicate this word-only approach to the bistable reversal task with an adequate sample size.

4.9 Conclusion

To conclude, in the present experiment, we identified an ERP effect in response to reversals of interpretation of ambiguous words which closely mirrored the time course and topography of the established Reversal Negativity for bistable visual figures. Whether the effect is one continuous RN, or two simultaneous components is as yet undetermined and hopefully future studies can shed light on this. We also identified a Late Positive Component in response to reversals of interpretation of ambiguous words which also closely mirrored the time course and topography of the established LPC for bistable visual figures. We interpret these findings as suggesting that more conceptual or interpretive types of ambiguities, such as those conveyed by ambiguous words, may exist as multiple stable representations in the brain that behave in bistable fashion similar to the multiple simultaneously possible representations of bistable visual figures. The ability to create coherence out of ambiguous sensory input is essential to our perception of the world. Our brains must constantly interpret an overwhelming but also limited and ambiguous amount of sensory information to generate our subjective experience of the world. This ability of the brain to modulate and resolve ambiguity is a crucial aspect of perception, language comprehension, and consciousness. It is one of the most wonderful and as yet almost unfathomable ways our conscious and unconscious experience is mediated by a series of chemical reactions.

Several theories of perception, language comprehension, and consciousness propose that the brain generates predictions and evaluates the reliability of stimulus inputs to construct our perceptions. For example, Predictive Coding Theory Rao & Ballard 1999) posits that the brain is constantly generating predictions and evaluating the reliability of sensory input in order to construct our perceptions. Similarly, constraint-based models of language comprehension (Prince & Smolensky 1993/2004) propose that the brain simultaneously processes multiple interpretations of a sentence to arrive at a coherent understanding. Higher-Order Thought theory suggests that consciousness itself may be dependent upon mechanisms of reality monitoring (Lau & Rosenthal 2011). The present findings suggest that complex information, such as conceptual representations of ambiguous words, may also behave in a bistable manner in the brain. This opens up the possibility that other levels of representation and abstraction may behave similarly and could provide insights into how the uncertain representations that are constantly being constructed in our minds, transition from ambiguity to subjective reality over time. If the present findings hold true with further research, they also shed light on what exactly the RN means in terms of the processing of bistable phenomena as it would appear the RN also exists in switching phenomena outside of visual perception, the framework in which the RN is usually understood.

Overall, these findings shed light on the brain's remarkable ability to process and make sense of ambiguous sensory input. They also highlight the potential for further research to uncover the mechanisms by which the brain resolves ambiguity and constructs our subjective experience of the world.

Bibliography

Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Perception & Psychophysics, 58*(2), 191–206. https://doi.org/10.3758/BF03211875

Bach, M., & Poloschek, C. M. (2006). Optical illusions. *Adv Clin Neurosci Rehabil, 6*(2), 20–21.

Bachmann, T., & Aru, J. (2023). Conscious interpretation: A distinct aspect for the neural markers of the contents of consciousness. *Consciousness and Cognition, 108*, 103471. https://doi.org/10.1016/j.concog.2023.103471

Bartsch, R. (1984). Norms, tolerance, lexical change, and context-dependence of meaning. *Journal of Pragmatics, 8*, 367-393. https://doi.org/10.1016/0378-2166(84)90029-8

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience, 11*(3), 235–260. https://doi.org/10.1162/089892999563373

Britz, J., & Pitts, M. A. (2011). Perceptual reversals during binocular rivalry: ERP components and their concomitant source differences. *Psychophysiology*, *48*(11), 1490–1499. https://doi.org/10.1111/j.1469-8986.2011.01222.x

Chen, L., Wassermann, D., Abrams, D. A., Kochalka, J., Gallardo-Diez, G., & Menon, V. (2019). The visual word form area (VWFA) is part of both language and attention circuitry. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-13634-z

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain: A Journal of Neurology, 125*(Pt 5), 1054–1069. https://doi.org/10.1093/brain/awf094

Cui, L., Wang, Y., Wang, H., Tian, S., & Kong, J. (2000). Human brain sub-systems for discrimination of visual shapes. *NeuroReport, 11*(11), 2415.

Davidson, G. D., & Pitts, M. A. (2014). Auditory event-related potentials associated with perceptual reversals of bistable pitch motion. *Frontiers in human neuroscience, 8*, 572. https://doi.org/10.3389/fnhum.2014.00572

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4*(7), 752–758. https://doi.org/10.1038/89551

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology, 41*(5), 665–678. https://doi.org/10.1111/j.1469-8986.2004.00193.x

Elston-Güttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language, 52,* 256–283. https://doi.org/10.1016/j.jml.2004.11.002.

Franklin, A., and M. Pitts. *"Laurel & Yanny": EEG Neural Correlates of an Auditory Bistable Language Stimulus.* Reed College Thesis (December 2021)

Glezer, L. S., Jiang, X., & Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the "visual word form area." *Neuron, 62*(2), 199–204. https://doi.org/10.1016/j.neuron.2009.03.017

Intaitė, M., Koivisto, M., Rukšėnas, O., & Revonsuo, A. (2010). Reversal negativity and bistable stimuli: Attention, awareness, or something else? *Brain and Cognition, 74*(1), 24–34.

Ishida, T. (2014). L2 semantic representation: Homograph norms of word association. *LET Journal of Central Japan, 26*, 13–22. https://doi.org/10.20656/letcj.26.0_13.

Ishida, T. (2019). The Effects of Meaning Dominance in the Time-Course of Activation of L2 Lexical Ambiguity Processing. *Journal of Psycholinguistic Research*, *48*(6), 1269–1284. https://doi.org/10.1007/s10936-019-09657-8

Kiss, G.R., Armstrong, C., Milroy, R., Piper, J. (1973). An associative thesaurus of English and its computer analysis. In: Aitken, A.J., Bailey, R.W., Hamilton-Smith, N. (eds.) *The Computer and Literary Studies*, University Press, Edinburgh

Kornmeier, J., & Bach, M. (2004). Early neural activity in Necker-cube reversal: Evidence for low-level processing of a gestalt phenomenon. *Psychophysiology*, *41*(1), 1–8.

Kornmeier, J., & Bach, M. (2006). Bistable perception — along the processing chain from ambiguous visual input to a stable percept. *International Journal of Psychophysiology*, *62*(2), 345–349.

Kornmeier, J., & Bach, M. (2012). Ambiguous Figures – What Happens in the Brain When Perception Changes But Not the Stimulus. *Frontiers in Human Neuroscience*, *6*.

Kornmeier, J., Ehm, W., Bigalke, H., & Bach, M. (2007). Discontinuous presentation of ambiguous figures: How interstimulus-interval durations affect reversal dynamics and ERPs. *Psychophysiology*, *44*(4), 552–560.

Kornmeier, J., & Bach, M. (2012). Ambiguous Figures – What Happens in the Brain When Perception Changes But Not the Stimulus. *Frontiers in Human Neuroscience, 6*, 51. https://doi.org/10.3389/fnhum.2012.00051

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences, 15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Lei, C., & Canseco-Gonzalez, E. (2020). *Neural signatures of perceptual and conceptual reversals* Reed College

Luck, S. J. (2014). An introduction to the event-related potential technique (Second edition). Cambridge, Massachusetts: *The MIT Press*.

Miki, K. (2012). How Japanese EFL learners access English homographic words: An analysis by a semantic relevance judgment task. *JACET Journal, 55*, 19–29

Nelson, D. L., McEvoy, C. L., Walling, J. R., & Wheeler, J. W. (1980). The University of South Florida homograph norms. *Behavior Research Methods & Instrumentation,* 12, 16–37. https://doi. org/10.3758/BF03208320. Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9,* 225–236. https://doi. org/10.3758/BF03196957.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*, 15–22. https://doi.org/10.1162/jocn.1992.4.1.15

Ortego, K., and E. Canseco-Gonzalez. *Is the Chicken Ready to Eat? Electrophysiological Signatures of Ambiguity in the Brain.* Reed College Thesis (May 2018)

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*, 280-291. https://doi.org/10.1016/j.cognition.2011.10.004

Pitts, M. A., Gavin, W. J., & Nerger, J. L. (2008). Early top-down influences on bistable perception revealed by event-related potentials. *Brain and Cognition*, *67*(1), 11–24.

Plaut, D. C., & Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. *Cognitive Neuropsychology, 28*(3–4), 251–275. https://doi.org/10.1080/02643294.2011.609812

Price, C. J., Price, C. J., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. S. J., & Friston, K. J. (1996). Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain, 119*(3), 919–931. https://doi.org/10.1093/brain/119.3.919

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. New York, NY. Technical Report, Rutgers University and

University of Colorado at Boulder, 1993. Rutgers Optimality Archive 537, 2002. Revised version published by Blackwell 2004.

Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences of the United States of America, 115,* 2323-2328. https://doi.org/10.1073/pnas.1714730115

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2(*1), Article 1. https://doi.org/10.1038/4580

Schaff, A. (1964). Unscharfe ausdrücke und die grenzen ihrer präzisierung. In Sprache und Erkenntnis: *Essays über die Philosophie der Sprache* (pp. 220-243). Rowohlt: Reinbek b/Hamburg.

Shen, Z.-Y., Tsai, Y.-T., & Lee, C.-L. (2015). Joint Influence of Metaphor Familiarity and Mental Imagery Ability on Action Metaphor Comprehension: An Event-Related Potential Study. *Language and Linguistics*, *16*(4), 615–637.

Snyder, J. S., Yerkes, B. D., & Pitts, M. A. (2015). Testing domain-general theories of perceptual awareness with auditory brain responses. *Trends in Cognitive Sciences*, *19*(6), 295–297.

Wang, Y., Kong, J., Tang, X., Zhuang, D., & Li, S. (2000). Event-related potential N270 is elicited by mental conflict processing in the human brain. *Neuroscience Letters, 293*(1), 17–20. https://doi.org/10.1016/S0304-3940(00)01480-4

Yang, H., Laforge, G., Stojanoski, B., Nichols, E. S., McRae, K., & Köhler, S. (2019). Late positive complex in event-related potentials tracks memory. signals when they are decision relevant. *Scientific reports*, *9*(1), 9469. https://doi.org/10.1038/s41598-019-45880-y

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., Croft, W., & Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 1766-1771. https://doi.org/10.1073/pnas.1520752113