Is the Chicken Ready to Eat?

Electrophysiological Signatures of Ambiguity in the Brain

A Thesis

Presented to

The Division of Philosophy, Religion, Psychology, and Linguistics

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Kevin Mark Ortego Jr.

May 2018

Approved for the Division

(Psychology)

Enriqueta Canseco-Gonzalez

Table of Contents

List of Figures

Abstract

The study of how the brain processes ambiguous visual stimuli has provided psychology with a wealth of information about how we form coherent representations of the world from inherently noisy and overwhelmingly dense sensory inputs. The Reversal Negativity (RN) is an event-related potential (ERP) elicited when one's subjective perception of a bistable ambiguous figure, such as the Necker Cube or Rat-Man drawing, switches from one of its possible interpretations to the other. The RN is thought to reflect a change in the perceptual configuration of a stimulus' current representation in the brain. The present study investigates whether ambiguous sentences having two valid interpretations (e.g. "The chicken is ready to eat.") are represented in a similar bistable fashion as these ambiguous figures. To investigate this question, we recorded brain activity in twelve participants while presenting ambiguous figures followed by disambiguated variants, and ambiguous sentences followed by line drawings depicting one of the sentence's two possible meanings. On each trial, participants indicated whether or not the disambiguating stimulus matched their subjective interpretation of the previously seen ambiguous figure or sentence. We then compared ERPs elicited by these disambiguating stimuli in mismatching (reversal) reports vs. matching (stable) reports. Replicating previous findings, we observed the typical RN associated with reversals of bistable visual figures. In response to reversals of our "bistable" ambiguous sentences, we identified a large, frontally-distributed negativity effect occurring over a similar time-course as the visual RN. 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 more abstract "conceptual" ambiguities such as those present when forming representations of sentences. We discuss possible alternative explanations and possible interpretations of this "conceptual" Reversal Negativity.

Chapter 1: Introduction

1.1 Why do we care about ambiguity?

The visual world we experience usually appears stable, consistent, and unambiguous, and we go through life reasonably convinced that what we see is congruent with the physical reality that presumably exists outside our minds. Sometimes, however, we encounter circumstances and stimuli that lead us to perceive things not as they really are. "Optical illusions," perhaps more accurately described as "visual" or "perceptual" illusions, considering that the optics of the eye have little to do with them, capture our fascination and challenge us to consider that what we see is not always what is really "out there" in the world. Visual illusions are often thought of as mistakes or malfunctionings of the brain, but are better considered as instances of the brain's normally perfectly good rules and computations being applied to unique situations where something out of the ordinary causes these computations to return unexpected results. These illusions offer psychologists a unique opportunity to get a glimpse of the brain's perceptual machinery in the gap where the physical details of a stimulus and the details we perceive in the mind's eye fail to overlap.

Visual illusions can allow us to dissociate the processes of sensation, in which our sense organs detect and transmit external stimuli 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 experience of the outside world, although where sensation ends and perception begins is not a well-defined line (Kornmeier et al 2011). In some cases, the dissociation between sensation and perception can be seen in the form of a perceptual artifact, something we perceive which is not present in the stimulus itself, such as the patches of darkness at the vertices of the Hermann Grid (Figure 1.1a) which disappear when we fixate on the vertices (Bach and Poloschek 2006). This illusion has traditionally been explained 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 sensing the physical light signal, although this explanation is incomplete and orientation selective neurons in the visual cortex are also likely at play. Another well-known illusion is the Zöllner illusion, in which parallel lines intersected repeatedly by shorter lines appear to diverge (Figure 1.1b). The mechanisms behind this illusion are still not fully understood, but this and similar geometric illusions seem to rely on an overestimation of acute angles, likely in the visual cortex. (Bach and Poloschek 2006).

Illusions such as the Hermann Grid and Zöllner illusion might be described as occurring due to “bottom-up” processes, those which are related more directly to sensory aspects of the stimuli themselves and their early processing, as opposed to “top-down” processes which are more cognitive, such as attention, expectancy, and decision making (Kornmeier and Bach 2012). Despite any mental effort on our part, we cannot make the gray patches disappear in the Hermann Grid or force Zöllner’s lines to properly align themselves in the mind’s eye. Another family of visual illusions that does however allow for top-down cognitive influences on our perception is the bistable figure, and it is these figures that form the basis of this thesis.

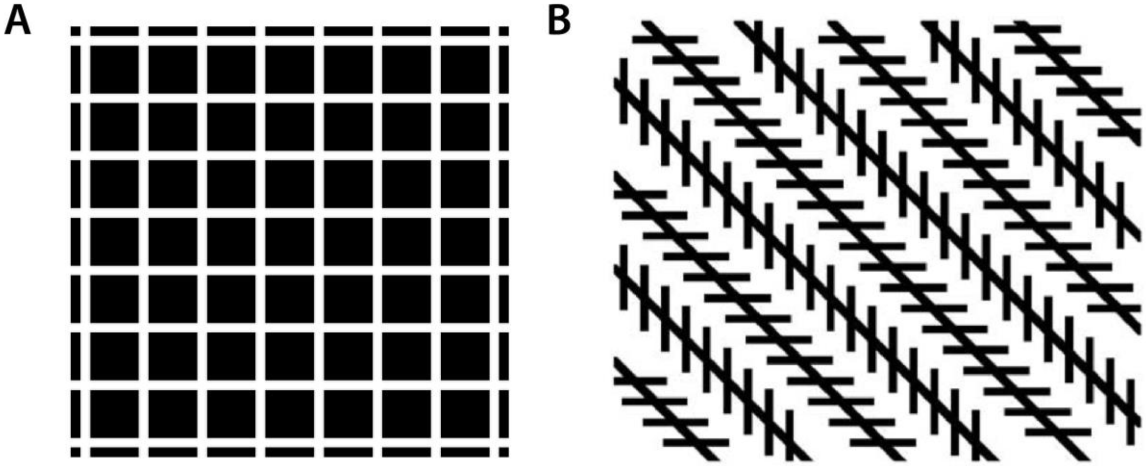


Figure 1.1 Bottom-Up Visual Illusions

(A) The Hermann Grid: one should perceive gray dots at the corners of each square, with any given dot disappearing if one fixates on its location. **(B)** Zöllner illusion: despite our perception, the large diagonal lines are actually parallel (overlaying a pencil between two of the lines makes this clear)

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, and Schroeder's Staircase (Figure 1.2). Despite the fact that the sensory input we receive at our eyes is constant, our brains readily switch between two interpretations of these images. Crucially, however, only one interpretation can be perceived at a time, and switches happen between these discrete interpretations, regardless of our knowledge that the identical information underlying both interpretations co-exists at the same time right in front of us.

I encourage the reader to spend a little time looking at these figures to appreciate the strangeness of this phenomenon (a personal favorite is the staircase). One can sometimes intentionally make these figures appear one way or the other. You can also intentionally prevent them from switching to the other interpretation for a while, but not forever, and they can never appear both ways at once. Or, you can do absolutely nothing but stare and watch as they switch seemingly on their own after a little time. We don't see anything that isn't on the page; there is no trickery to be exposed with the use of a ruler. It's a shift in the mind and only in the mind that you can feel viscerally, and it's very bizarre and very cool indeed.

The fact that these discrete and well-defined reversals of perception occur in response to viewing bistable figures has made them a subject of research interest, particularly in regards to the neural activity that underlies these discrete perceptual reversals of an identical physical stimulus. One technique that allows us to examine activity in the brain in the specific moments surrounding perceptual reversals is the recording of brain electrical activity via EEG.

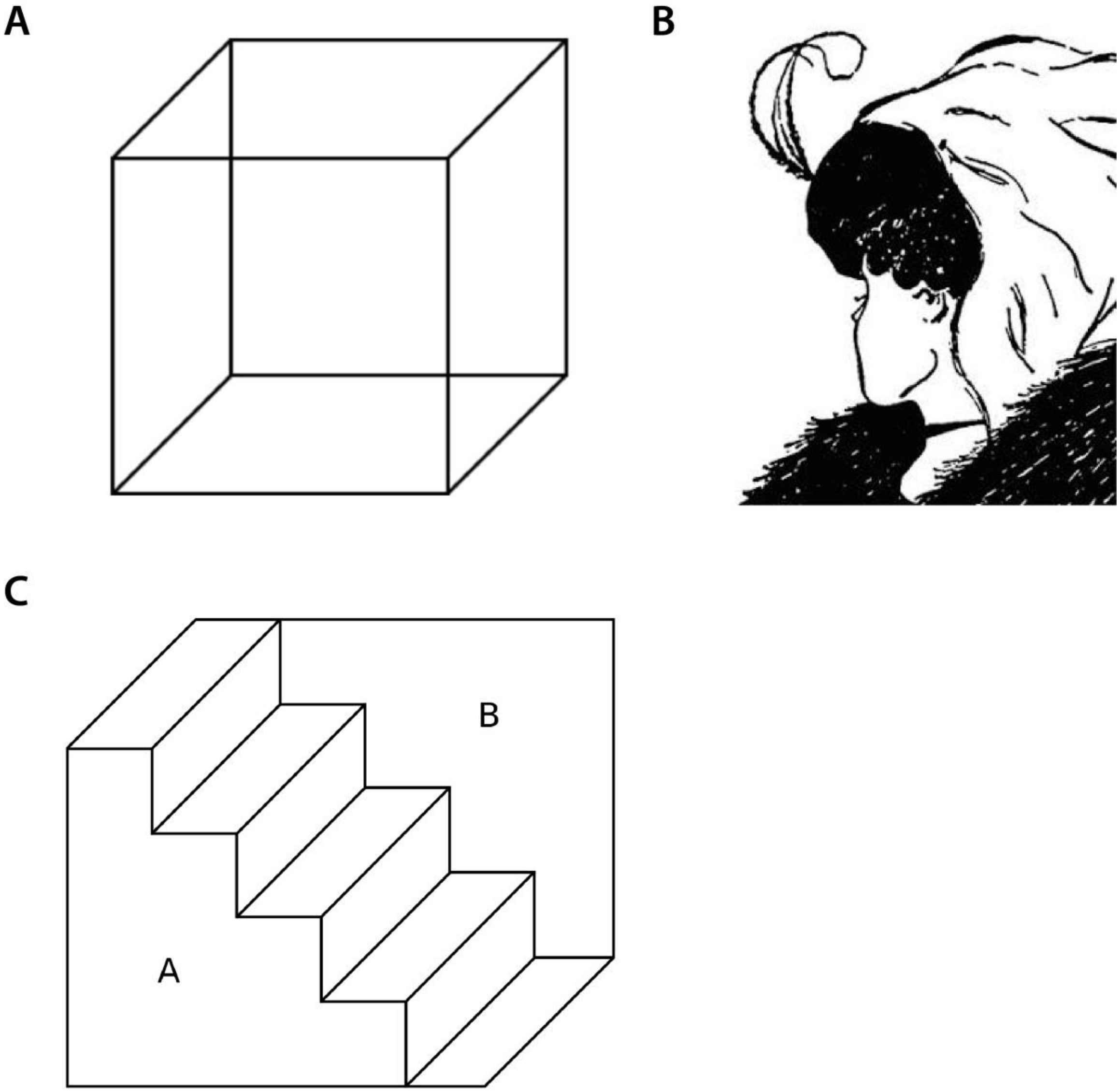


Figure 1.2 Common Bistable Figures

(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. **(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 an upside-down staircase hanging from the "ceiling" of the image with region B in the foreground. If you're having trouble seeing the latter possibility, just keep staring at the center of the image and it should appear after several seconds.

1.3 What is EEG and what are ERPs?

Electroencephalography (EEG) is a method of recording brain electrical activity using an array of electrodes placed on the scalp. When a neuron fires, an electrochemical action potential is propagated along the axon to terminal endings where neurotransmitters are released into the synapse between neurons. These transmitters then bind to receptors on the dendrites of the postsynaptic neuron, generating another electrochemical potential, this time called the postsynaptic potential. Action potentials traveling along axons are extremely brief, lasting perhaps only one millisecond, whereas postsynaptic potentials can last from tens to hundreds of milliseconds. Due to the organization of neurons in the cortex and their short duration, action potentials do not summate strongly enough to be detected at the scalp. Postsynaptic potentials, however, occurring in the dendrites and cell bodies which tend to be oriented similarly in cortical tissue, can and do, and it is these potentials that we record using EEG (Luck 2014).

At any given moment, neurons across the brain are spontaneously firing, essentially generating random noise, but when a cognitive task is performed or a stimulus is observed, coordinated activity in certain brain regions will result. By recording EEG while presenting many trials and “time-locking” the resulting data to some event in these trials, such as the appearance of a stimulus or a button press to indicate a response, random noise tends to cancel itself out after the averaging of many trials. This leaves only neuronal activity related to the event itself, thus the name event-related potential. By comparing ERPs elicited by different experimental conditions, we can examine differences in brain activity between conditions on a time scale of milliseconds. Some features of ERP waveforms reliably result in response to specific tasks or types of stimuli, such as making a decision about whether to respond to a stimulus or hearing a deviant tone in a stream of consistent tones. These consistently observed sub-parts of waveforms are referred to as components and often receive names that reflect their voltage, timescale, or function (Luck 2014).

1.4 Electrophysiological Correlates of Bistable Perception

1.4.1 The Reversal Negativity

The Reversal Negativity (RN) is one ERP component that has been found in response to the viewing of bistable figures (Kornmeier and Bach 2004). When bistable figures are presented briefly and sequentially with a short interval of time between presentations, one's subjective interpretation of the figures will periodically and spontaneously switch from trial to trial. To model this effect, you can return to Figure 1.2 and repeatedly open and close your eyes at roughly half second intervals while fixating in the center of each figure. Every few blinks, your interpretation of the figure should switch upon opening your eyes. The Reversal Negativity is the ERP signature obtained when comparing trials in which your interpretation of the figure switches relative to the previous trial, to trials in which perception remained stable compared to the previous trial. Specifically, the RN is computed by taking the averaged ERPs in response to trials in which a reversal of perception occurs and subtracting from it the averaged ERPs of trials in which perception remained constant. Practically speaking, when you blink and open your eyes and the figure remains the same as before, something different happens in the brain compared to when you blink and find the figure reversed, and this is what is reflected in the RN.

As a control condition in these experiments, researchers have constructed disambiguated versions of bistable figures and presented these sequentially to compare the effects of endogenous (i.e. spontaneous) switches between the bistable percepts and exogenous switches triggered by a physical change in the stimulus. The ERPs in response to these control conditions reveal that the brain does in fact respond differently to stimulus-driven exogenous reversals and percept-driven endogenous reversals. A Reversal Negativity is still found for exogenous reversals, but it peaks around 220ms after stimulus onset, as compared to 260ms after onset for endogenous reversals (Kornmeier and Bach 2006). Additionally, a Reversal Positivity (RP) occurring around 130ms after stimulus onset has been observed in response to endogenous reversals, but not in response to exogenous reversals of disambiguated stimuli. (Kornmeier and Bach 2012). Kornmeier and Bach (2011) consider the RP to be reflective of an initial detection of stimulus ambiguity via processing conflicts 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 variants.

The causes and functional role of the Reversal Negativity are still not fully understood, considering that it is subject to both bottom-up and top-down influences and occurs with both endogenous and exogenous reversals, and proposed explanations include that it may be reflective in a general sense to a change in the perceptual configuration of a representation in the brain (Intaite et al. 2010, Kornmeier and Bach 2012). Pitts et al. (2008) found that the amplitude of the RN was increased when participants were instructed to intentionally cause reversals of interpretation as frequently as possible, as compared to passively viewing the figures and waiting for spontaneous reversals, which suggests that top-down control can modulate the processing of bistable figures as early as 150ms after image onset. While clearly susceptible to top-down influence, bottom-up factors such as duration of stimulus presentation and duration of the inter-stimulus interval between presentations also modulate reversals, with reversal rates being maximized with an inter-stimulus interval of around 400ms and decreasing with longer intervals (Kornmeier et al. 2007). Kornmeier et al. (2009) found that the effects of top-down intentions to cause reversals were additive with the effect of shortening inter-stimulus interval, suggesting that perceptual reversals are susceptible to the influence of multiple neural mechanisms simultaneously.

Intaite et al. (2010) investigated the RN's relationship to the attention-related N2pc, an ERP component which occurs in posterior scalp regions contralateral to a presented stimulus, by using a bilateral display of two Necker lattices and examining whether an N2pc occurred when one of the two lattices reversed. The authors concluded that the RN was not a variant of the N2pc and thus not caused solely by attentional effects, and additionally note that the RN was not correlated with the specific variants of the Necker lattices in subjective awareness. They interpret this result as suggesting that RN thus reflects a general change in the contents of perceptual awareness. Kornmeier and Bach (2012) offer a similar interpretation framed in terms of switches between the multiple perceptual “attractor” states which could correspond to the current ambiguous visual stimulus. They postulate that an unambiguous visual stimulus generates a single powerful attractor, but that these bistable figures may generate multiple less-stable attractors close to one another in perceptual space, with reversals occurring 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.

1.4.2 The Late-Positive Component

The Late-Positive Component (LPC) is another ERP component reliably elicited in studies of perceptual reversals of bistable figures, and occurs in response to both endogenous and exogenous reversals (Kornmeier and Bach 2006). Like the RN, the LPC is a difference obtained by comparing waveforms of reversal trials to stable trials, and manifests as an enhanced positivity for reversal trials compared to stable trials, beginning approximately ~350ms after stimulus onset. The LPC is thought to reflect the updating of the contents of visual short-term working memory to account for the reversal of the bistable stimulus (Pitts et al. 2007). In the context of typical reversal paradigms, the RN may reflect the shift in the current perceptual configuration of a bistable stimulus, and the LPC may reflect the encoding of this new percept into memory in order to make perceptual comparisons on subsequent trials and to indicate with a response that perception has changed (Pitts et al. 2009). As such, the LPC may be intrinsically linked to task-demands of the requirement to report reversals in these paradigms, rather than being a signature of the perceptual reversal itself.

1.5 Ambiguity in Other Domains

The Reversal Negativity and Late-Positive Component effects have also been observed in response to perceptual shifts in binocular rivalry paradigms, where competing stimuli are presented monocularly to each eye and perception spontaneously alternates between the left-eye or right-eye stimulus (Britz and Pitts 2011). One could consider the two alternative percepts in binocular rivalry to behave in a bistable fashion, and the phenomenon of bistable perception itself is not unique to vision. 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). Bistable perception can even occur in the olfactory system when presenting different odorants to each of the two nostrils, with one's perceived smell switching back and forth in a similar manner once again (Zhou and Chen 2009). Time-locking EEG to subjective reversals of auditory and olfactory percepts is logistically more difficult than time-locking to the onset of a visual stimulus, and to the best of our knowledge no studies exploring a potential RN in auditory stream segregation or olfactory rivalry have been conducted. An auditory analogue of the RN has been identified, however, in response to sequentially presented complex tones which can be bistably perceived as ascending or descending in pitch, with discrete reversals happening only at tone onset (Davidson and Pitts 2014). Another domain that is ripe with ambiguity and which may be lend itself to exploration of bistability with EEG is that of language processing, and the possibility of bistable linguistic stimuli and a corresponding Reversal Negativity in language is the focus of this study.

1.6 Ambiguity in Language

Ambiguity is present at all levels of language processing, from the perception of speech sounds themselves to the meanings of sentences, and our brains constantly work to categorize and interpret often noisy linguistic stimuli into comprehensible language. A fun illustration of the brain's attempts to deal with ambiguous information in language processing at the sound level is the McGurk effect (McGurk and MacDonald 1976). When video of an individual mouthing the syllable /ga/ is presented with mismatched audio, such as the sound /ba/, our brains tend to combine the two streams of conflicting information into a percept somewhere in between, in this case perceiving the syllable /da/, even though this syllable matches neither the "true" auditory nor "true" visual inputs. This effect occurs for a range of other syllable combinations, and is a powerful example of how the brain can synthesize information from multiple sensory modalities in order to create a stable percept out of an ambiguous input.

The McGurk effect and its reliance on cross-modal integration to resolve ambiguity (which results in an illusory intermediate percept) is ultimately a substantially different phenomenon from bistable perception. Oronyms, however, are phrases that sound similar to one another, but which can be parsed in two different ways, such as the sentences "The stuffy nose can lead to problems," and "The stuff he knows can lead to problems." Here, there are two valid interpretations of an identical auditory input, and one could presumably switch between these two interpretations with repeated presentations of the sentences. Additionally, many oronyms rely on unconventional pronunciations to create ambiguity and therefore lend themselves to disambiguation via subtle differences in emphasis in speech, but here again we are faced with the difficulty of time-locking to the precise point of perceptual shifts with auditory stimuli. One could theoretically present an oronym auditorily and then potentially trigger a "reversal" of interpretation via presentation of an image corresponding to one of its two interpretations, but many oronyms are heavily biased toward one more sensible interpretation due to their meanings being constrained by the requirement for phonetic similarity, as in the pair "Peace talks were needed to prevent war," and "Pea stalks were needed to prevent war." Oronyms, because of their reliance on auditory ambiguity are not the best avenue for exploring linguistic bistability, and the written word may provide a better avenue for study with ERPs.

Because they require reading or hearing a sequence of words over time, ambiguity is also present during the active construction of the meanings of sentences, and many of these ambiguities are temporary in nature and can be swiftly resolved with later information. When reading the sentences "The old train the young" or "The old train left the station," we encounter the same three initial words, "the old train," which could refer either to old people performing the action of training, or to an old locomotive. The momentary ambiguity upon seeing or hearing these first three words is resolved so quickly when we encounter the rest of the sentence that we scarcely notice it was ever there. In other sentences, however, later information can momentarily render what we have already read more unclear rather than less. This is the case for "garden-path" sentences, so called because they lead the reader down a predictable and attractive path of understanding until some unexpected element comes along that no longer makes sense, triggering a complete reinterpretation of the sentence. A classic example is "the horse raced past the barn fell." The initial interpretation of this sentence involves a horse galloping past a barn, and the trouble arises when we encounter the seemingly out of place "fell," which ends the sentence. The correct interpretation of the sentence is of a horse being raced, presumably by some unnamed rider, which then falls after passing the barn. This ambiguity is the result of using a reduced relative clause (one that lacks a relative pronoun) and could be avoided altogether by including an appropriate relative pronoun, as in "The horse *that* raced past the barn fell," which now clearly specifies that we are talking about a specific horse which raced past a barn, then fell.

These types of ambiguities in the domain of the active processing of language have been studied extensively, including with ERPs, but they are not ambiguous in the same sense that bistable figures are, because ultimately there is a single most-favorable interpretation that the brain will settle upon once all the necessary information is available and integrated. There are however, linguistic stimuli which we believe could be considered analogous to bistable figures, and they result when ambiguities within a sentence cannot be resolved by context or the likelihood of one interpretation, even after all the necessary information is available to the brain.

1.7 Linguistic Analogues of Bistable Figures

When presented with the sentence, "The chicken is ready to eat," you likely envision a nicely roasted chicken coming hot out the oven, or perhaps instead you think of a chicken pecking at some feed that a farmer has poured in the coop. Maybe a vegetarian would first settle on the latter interpretation, whereas someone who's hungry would arrive at the former. Although personal biases may influence which interpretation occurs first, both are equally valid and plausible readings in the absence of any disambiguating context. The noun "chicken" can refer both to the meat of the animal and to the living animal, and interestingly enough, this example would not work with cows or pigs because the meats of these animals are signified by different nouns in English (beef and pork respectively). Thus the ambiguity here lies in the fact that the word "chicken" has multiple meanings, each of which fits the overall meaning of the sentence. Ambiguities that result from a word's multiple meanings are referred to as lexical ambiguities, and interestingly, one's resolution of the lexical ambiguity in a sentence can cause the entire structure of the sentence to change. In "the chicken is ready to eat," when we interpret "chicken" as the living animal, "the chicken" is then interpreted as the agent of the sentence, the "doer" of the verb "eat". In contrast, if "the chicken" is interpreted as food, then the same phrase plays the role of the object of the verb "eat." Here, the meaning of a single word can reverse our entire conceptualization of the sentence. A similar lexical ambiguity is present in the sentence "I saw her duck," in which the word duck can be interpreted either as a noun naming an animal owned by some woman, or as a verb, signifying the action that the woman is performing, and again, the ambiguity of this single word causes a complete restructuring of our understanding of the sentence.

Another type of ambiguity that results in multiple valid interpretations is anaphoric ambiguity, a type of syntactic ambiguity in which a phrase could plausibly attach to multiple elements in the sentence. As an example, "The woman hit the man with the umbrella," can be interpreted as a woman using an umbrella to hit a man, or as a woman hitting a man who is holding an umbrella. This is due to the fact that the prepositional phrase "with the umbrella" can modify either the manner of the woman's action of hitting the man, or modify the noun phrase "the man" to specify which particular man we are talking about, with both possibilities being equally plausible.

Ambiguities such as these can often be disambiguated by the context in which they occur. In a story describing a dinner party, "the chicken is ready to eat" would clearly refer to the chicken-as-food possibility, whereas a story about children at a petting zoo would lead to the opposite interpretation. Similarly, if it were intended for “the woman hit the man with the umbrella” to refer to a woman *using an umbrella* to hit a man, we’d presumably have received some prior information that lets us know that this man deserves to be hit, or if the other meaning was intended, we’d have been made aware that there are multiple men who could possibly get hit by this woman, with the phrase “with the umbrella” serving to identify her victim.

Somewhat like the debate over the relative contribution of bottom-up and top-down processes in the perception of ambiguous figures, psycholinguists debate the extent to which the brain relies strictly on incoming linguistic input to generate representations of sentences, versus the extent to which multiple plausible representations of meaning are constructed in parallel before a final alternative is settled upon (Sedivy 2014). The “garden path” theory proposed by Frazier and Fodor (1978) holds that the brain constructs a single most likely interpretation of a sentence while reading, which is only reevaluated if later information conflicts with the current most likely interpretation. This could be considered a more “bottom-up” approach to sentence parsing, relying exclusively on the incoming linguistic stimulus. The competing “constraint-based” approach of MacDonald et al. (1994) proposes that multiple possible interpretations of a sentence are constructed in parallel so long as they are allowable according to the present constraints of the sentence, and that the multiple possibilities are pruned away and a final interpretation decided upon as new information and new constraints are revealed with further reading. This model, with its reliance on expectancy and broader contexts, might be considered the analogue of more “top-down” processing.

Both of these models offer valuable insight for understanding why ambiguities during active sentence reading, such as those caused by reduced relative clauses in so-called “garden path” sentences, may cause us so much trouble in some cases (“The horse raced past the barn fell.”) but not in other cases using reduced relative clauses, such as “The treasure buried in the sand was never found,” (see Sedivy 2014 for discussion of these models and their interpretations). Critically, however, neither of these models is capable of definitively resolving the ambiguities present in sentences such as “The chicken is ready to eat,” “I saw her duck, or “The woman hit the man with the umbrella.”

While a reader will settle on one or the other interpretation of these sentences upon an initial reading, possibly determined by some cognitively biasing factor such as hunger level or vegetarianism in the case of the chicken, once the ambiguities are realized, there is no information in the sentences themselves that would allow us to determine which interpretation is "correct" in the absence of any external context. In the same way that a bistable figure delivers a complete set of visual information to the brain that can be interpreted in one of two mutually exclusive but equally valid ways, these ambiguous sentences could be considered complete linguistic stimuli with two mutually exclusive and equally valid possible interpretations, and we therefore believe that these fully ambiguous sentences are a plausible linguistic analogue of bistable visual figures.

1.8 Rationale and Hypotheses

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 and visual ambiguities, or whether analogous processes are involved as part of a more general perceptual switching phenomenon when confronted with ambiguity, in this case, in the domain of language when processing fully ambiguous sentences. In order to investigate the reversal negativity in the linguistic domain, we will repeatedly present fully ambiguous sentences, such as those discussed above, which will be then be disambiguated by presentation of line drawings depicting one of the two disambiguated interpretations of the sentence. By achieving disambiguation using visual stimuli, we can time-lock ERPs to the onset of the disambiguating image for analysis. Visual stimuli provide the further advantage of delivering the necessary disambiguating information all at once, rather than having to provide disambiguating context in words, thus avoiding the problem of variations in reading times making it difficult to time lock ERPs for analysis.

The necessity of presenting an ambiguous sentence first followed by a disambiguating image in our paradigm contrasts with previous studies of bistable figures in which stimuli were exclusively ambiguous, exclusively unambiguous, or consisted of an initial unambiguous stimulus followed by an ambiguous variant (see Intaite et al.

2013). Because our ambiguous-then-unambiguous method of presentation is novel, it will first be necessary to employ this presentation sequence using bistable visual figures in order to confirm that the RN and LPC are still observed when stimuli are presented in this order.

If we successfully replicate the Reversal Negativity and Late-Positive Component effects for visual figures using our novel presentation paradigm, and if these ambiguous sentences behave similarly to bistable figures in the brain, we would expect to find ERP signatures similar to the RN and LPC when comparing trials in which the disambiguating image mismatches with a participant's interpretation of the sentence versus trials in which the disambiguating sentence matches the participant's interpretation. Given that language is processed in different areas of the brain than vision and that the meanings of sentences are likely represented in a different manner than visual perceptions corresponding to a physical stimulus, it is likely that the location and time course of any observed effect in response to reversals of these sentences may vary from those observed in traditional bistable figures paradigms.

1.8.1 Previous Similar Paradigms and Possible Predictions

Interactions between visual stimuli and language processing have previously been demonstrated, although not in a paradigm identical to this one to the best of our knowledge. The most similar experiments to ours include those using picture-sentence verification tasks, in which a picture is presented to establish a context prior to the presentation of a sentence which may or may not deliver a meaning congruent to the established context. Knoeferle et al. (2011) presented pictures followed by sentences which described events either congruent or incongruent with the images. For example, an image was presented of a gymnast either punching or applauding a journalist, which was followed either by the sentence "The gymnast punched the journalist" or by "The gymnast applauded the journalist". These sentences were visually presented word-by-word to allow ERPs to be time locked to the critical verb ("punched" or "applauded") that determined whether the sentence matched or not.

When the critical verb was presented, the authors found a large N400 effect for the verbs that mismatched as compared to those that matched. The N400 is an ERP component that occurs when violations of semantic expectancies occur, such would happen in the sentence “I take my coffee with cream and *dog*” when the word “dog” is read. The N400 is also sensitive to semantic relatedness in a more general sense. It has been observed when presenting pairs of unrelated versus related words in isolation, and has also been found when presenting sentences word-by-word when the critical word is presented as an image (Nigam 1992; and see Kutas and Federmeier 2012 for an overview of the N400 literature). Therefore, it is possible that we may observe some type of N400-like effect in our study in response to the disambiguating image violating the participants' semantic expectancies based on the interpretation of the sentences that they form.

Another language-related ERP component potentially relevant to the present study is the P600. The P600 is traditionally considered to occur in response to perceived syntactic violations, such as those which occur in garden path sentences like “The horse raced past the barn fell,” as discussed earlier. The P600 has also been observed in syntactically unambiguous sentences in response to some types of semantic anomalies, such as in the sentence “the javelin has thrown the athletes,” (Vissers et al. 2007). In another picture-sentence matching experiment Vissers et al. (2007) found a large P600 effect in response to sentences which inaccurately described the relationship between a dyad of previously presented shapes, and the authors make a case that the P600 may reflect the initiation of a complete reprocessing of a sentence to check for correctness of an interpretation, rather than resulting from syntactic violations specifically.

The evidence from these paradigms establishes that visual information does interact with linguistic processing in the brain in ways that can be measured using ERPs. Unlike our experiment, however, the sentences used in these experiments were unambiguous, and these sentences were presented word-by-word after pictures for comparison with the pictures, rather than presenting a sentence followed by a picture. It is unclear whether components such as the N400 and P600, which are normally investigated and observed in the context of active construction of a sentence's meaning, would be expected to occur in response to the disambiguation of an already completed and stable interpretation of an ambiguous sentence. The appearance of these components seems plausible, however, given the results of these picture matching tasks, and their possible occurrence does not rule out the possibility of also observing RN or LPC-like effects.

Chapter 2: Methods

2.1 Participants

A total of 12 current Reed College students (6 female; mean age = 20.9) with normal or corrected-to-normal vision and no history of brain injury participated in this study. Compensation for participation included eight tickets for the chance to win a $50 prize in the Reed College Psychology Department lottery. All procedures were approved by the Reed College Institutional Review Board.

2.2 Stimuli

Ambiguous figure stimuli (Figure 2.1) consisted of ambiguous and disambiguated versions of a Necker Lattice, a modified Schroeder staircase, and Bugelski and Alampay's Rat-Man illusion. Ambiguous sentence stimuli consisted of the sentences "The chicken is ready to eat," "I saw her duck," and "She hit the man with the umbrella." Each sentence was paired with two disambiguating illustrations, each corresponding to one possible interpretation of the sentence and roughly matched for visual complexity (Figure 2.2). Necker cube and Schroeder staircase stimuli were created using Adobe Illustrator. Rat-Man stimuli were taken from a previous thesis project (Jimenez 2017). Disambiguating sentence illustrations were hand-drawn and scanned as high-resolution JPEG files. All stimuli were presented using Presentation (Neurobehavioral Systems, Berkeley CA).

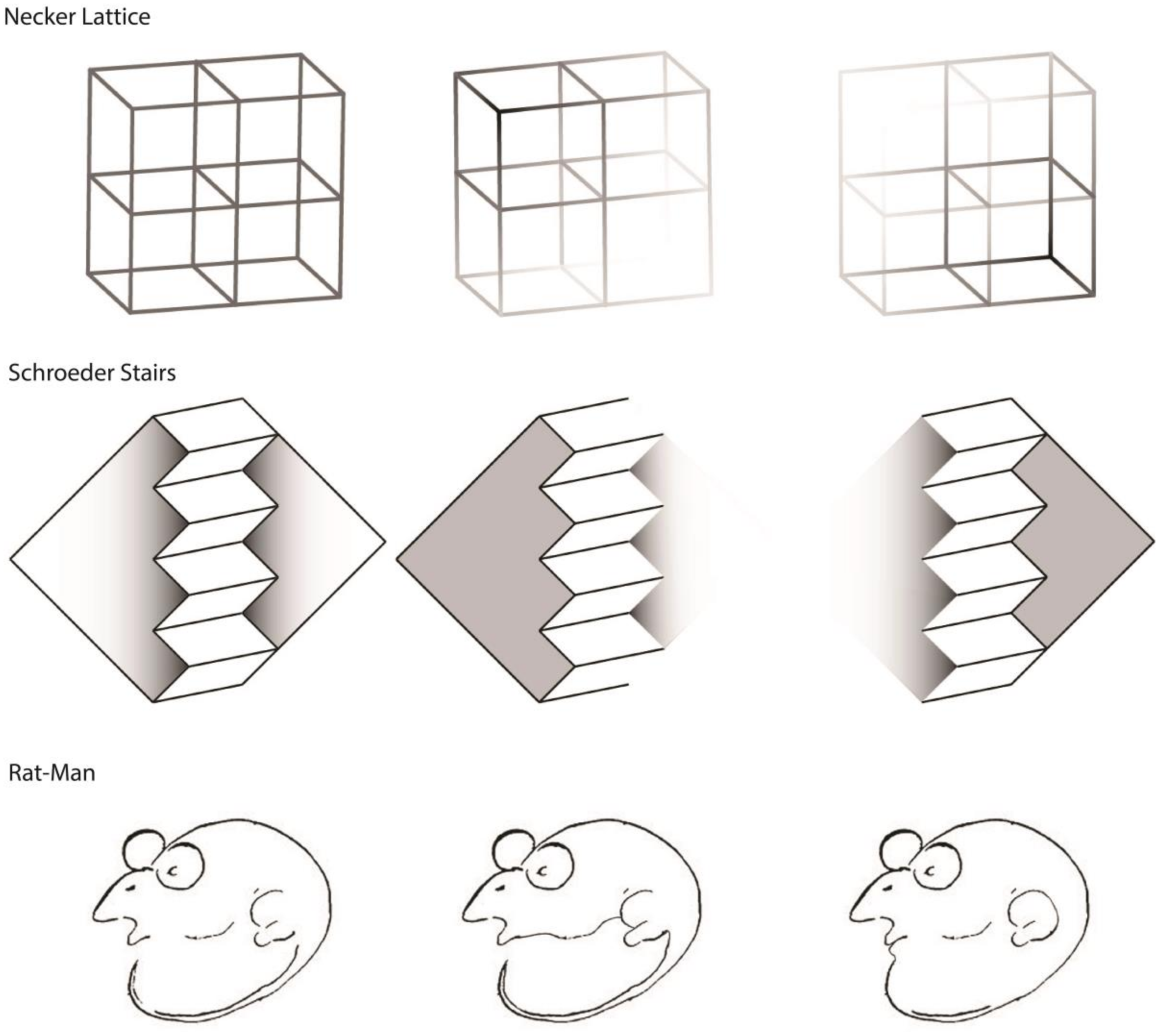
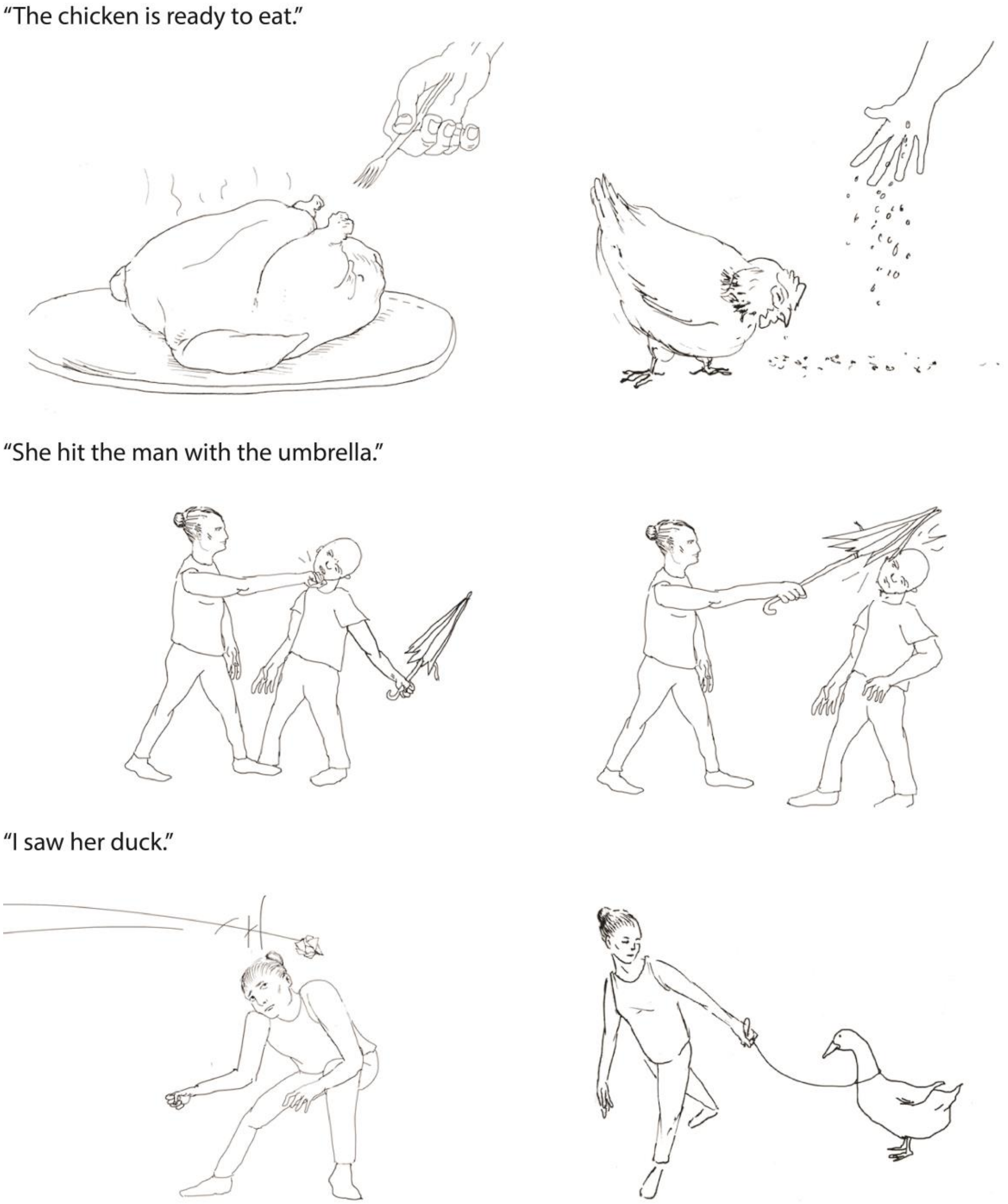


Figure 2.1 Visual Condition Stimuli
Ambiguous visual figures on the left column and their two disambiguated variants in the middle and right columns.



"The chicken is ready to eat."

"She hit the man with the umbrella."

"I saw her duck."

Figure 2.2 Sentences Condition Stimuli

Ambiguous sentences and the illustrations of their two disambiguated interpretations.

2.3 Procedure

All recordings took place in an electrically-shielded and sound-attenuated recording chamber with subjects seated ~72 cm from the screen. After the electrode cap was fitted, the experimenter guided participants through a familiarization block to ensure that they were aware of both possible interpretations of the ambiguous figures and sentences. If a participant was unable to perceive both interpretations of the figure stimuli, the experimenter provided cues by tracing the outline of the alternative percept on the screen. Participants were free to examine the stimuli for as long as they felt necessary to establish ease of switching between interpretations. After the familiarization phase, participants were given twenty practice trials with each stimulus in order to gain familiarity with the task.

The main experiment consisted of two rounds of six blocks, with each round consisting of three blocks of visual figure stimuli followed by three blocks of sentence stimuli, or vice versa, with the order of figure and sentence conditions counterbalanced across participants. Each block contained 180 trials, with a single ambiguous figure or sentence and its disambiguating images being presented in each block. Participants were given short self-paced breaks every ten trials and longer experimenter-controlled breaks every 60 trials.

In each figure condition, the ambiguous stimulus was presented for 500ms, followed by a 400ms delay after which the unambiguous figure was presented for 500ms. Following another 400ms delay, the participant was prompted to indicate whether their subjective perception of the ambiguous figure matched the orientation of the unambiguous figure. Participants responded “Match” or “Mismatch” using their right index finger to press one of two buttons (respectively) on a button box, and the next trial began automatically after participant response (Figure 2.3. Procedural details were identical for the sentences condition, except that the ambiguous sentences were presented for 800ms to allow for adequate reading time. (Figures 2.4)

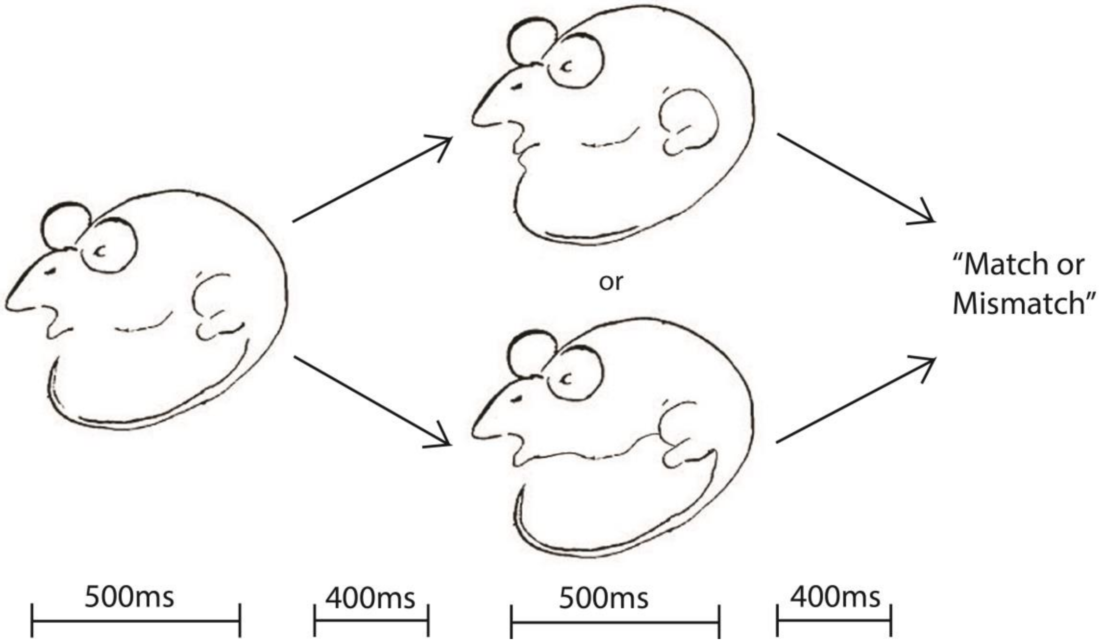


Figure 2.3 Diagram of a Visual Condition Trial

Each ambiguous visual stimulus was presented for 500ms during which participants perceived one of the two alternative percepts of the figure, followed by a 400ms blank screen before one of the unambiguous variants was presented. Following another 400ms blank, participants were prompted to indicate whether the unambiguous figure matched or mismatched with their subjective percept of the first figure.

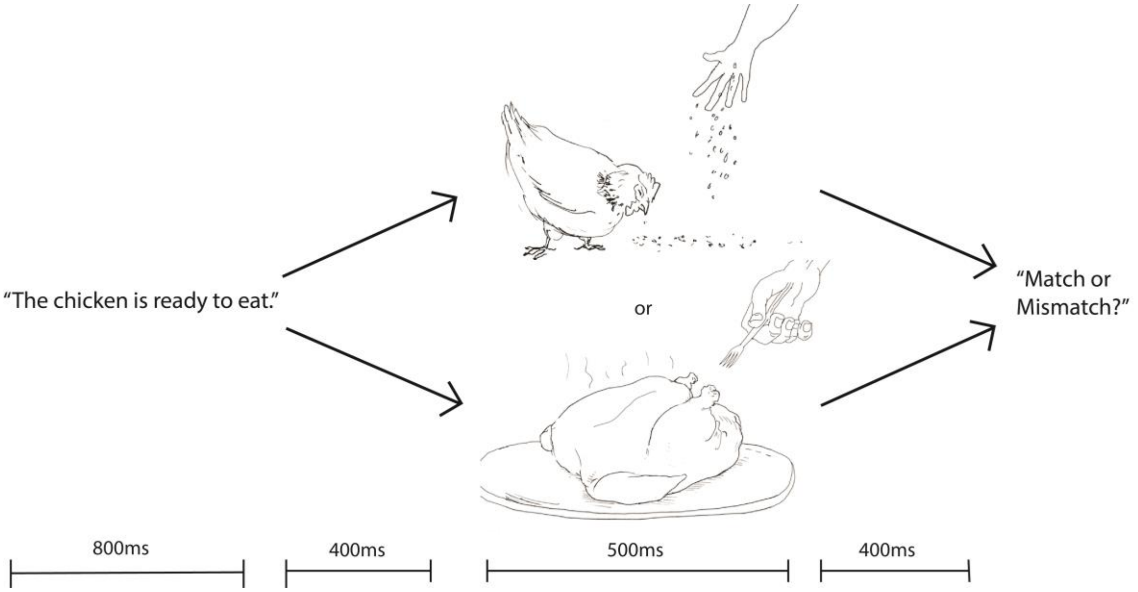


Figure 2.4 Diagram of a Sentences Condition Trial
Each ambiguous sentence was presented for 800ms during which participants read and formed a stable interpretation of the sentence, followed by a 400ms blank screen before one of the disambiguating images was presented. After another 400ms blank, participants were prompted to indicate whether the image presented matched or mismatched their interpretation of the sentence.

2.4 EEG Recording

In the EEG recording sessions, participants were fitted with a 64-channel electrode cap (Figure 2.5). An electrode placed on the face below the left eye (VEOG) was used to detect eye-blink artifacts, and two electrodes adjacent to the left and right eyes (HEOG) were used to detect horizontal eye movements. 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 4 hours of participants' arrival, caps were removed and participants were able to wash their hair in the lab.

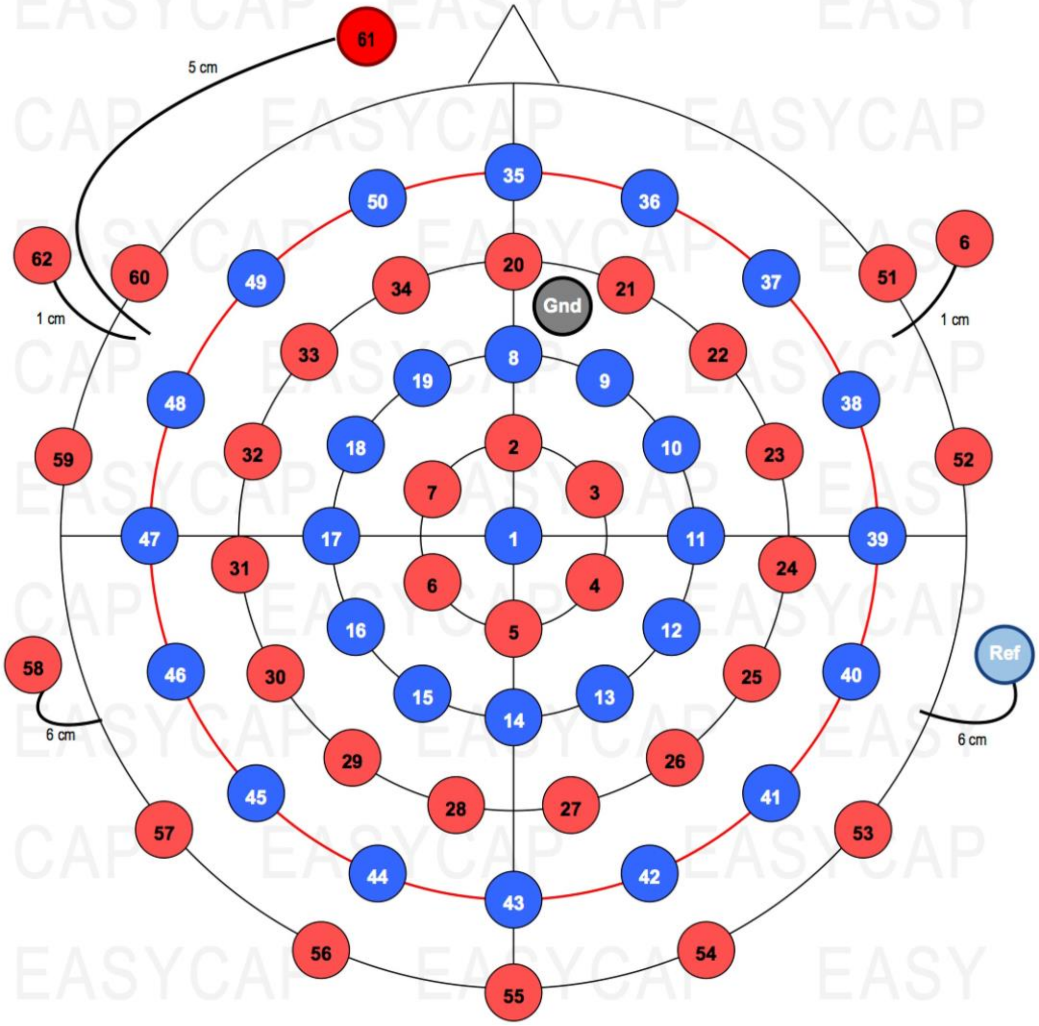


Figure 2.5 Electrode Locations
Diagram of the 64-channel electrode cap used in the experiment. Channel 61 corresponds to VEOG and channels 62 and 63 correspond to HEOG. Channels 58 and "Ref" correspond to the left and right mastoids respectively.

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 21% 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 mastoid electrodes. ERPs in each condition were time locked to onset of the disambiguating stimuli for all analyses.

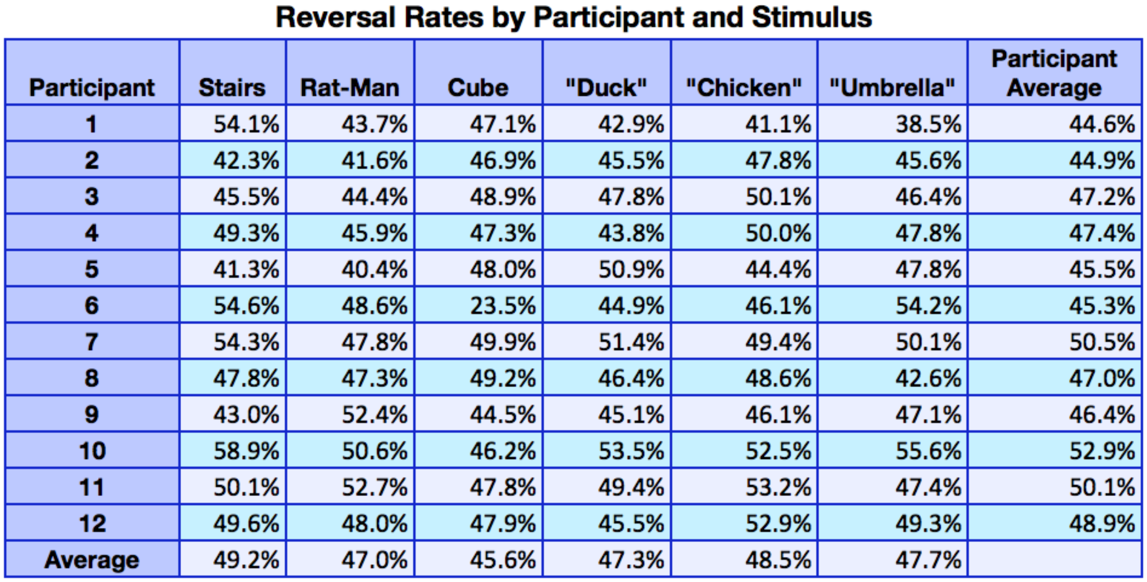
Based on our hypotheses and the results of previous studies, we measured the mean amplitude in two time windows in the figures condition. We were primarily interested in the reversal negativity (RN), which is a negative-going difference that occurs in right-posterior electrode sites between ~150-350ms on trials in which there is a perceptual reversal as compared to trials where perception remains stable (Pitts et al. 2008). Additionally, we investigated the presence of the late positive component (LPC), which is a positive-going difference that begins at ~350ms in central electrode sites in trials in which there is a reversal of perception as compared to stable trials (Pitts et al. 2008).

In the sentences condition, we first tested for the presence of RN and LPC effects similar to those of the visual condition in their respective time windows and regions. Given the exploratory nature of the sentences condition, we then visually inspected the waveforms for the presence of language-related ERP components such as the N400 and P600, or other apparent effects on which to perform analysis. No waveforms resembling an N400 or P600 were identified. However, a large negative-going difference occurring in a similar time-window (~150-350ms) to the RN on reversal trials as compared to stable trials was identified in frontal electrode sites. Based on examination of scalp maps of this component, we defined a left-frontal and right-frontal region of interest in which to compute mean amplitudes during the same time window as the RN.

Chapter 3: Results

3.1 Behavioral Results

Mean percentage of reversal trials across all stimuli and conditions was 47.5%, and mean percentage of reversals for each stimulus by participant are presented in Figure 3.1 below. Percentages of reversals in the visual figures condition (M=47.1% SD=2.9%) did not differ from percentages of reversals in the sentences condition (M=47.8% SD=3.1%; t(11) = .64, ns). Percentages of trials rejected for artifacts in the visual figures condition (M=16.3% SD=8.8%) did not differ from the sentences condition (M=26.4% SD=18.0%; t(11) = 1.82, ns). Additionally, percentages of reversal trials available for analysis after artifact rejection in the visual figures condition (M=47.0% SD=3.0%) did not differ from the sentences condition (M=47.7% SD=2.9%; t(11) = .89, ns).



Reversal Rates by Participant and Stimulus

Participant	Stairs	Rat-Man	Cube	"Duck"	"Chicken"	"Umbrella"	Participant Average
1	54.1%	43.7%	47.1%	42.9%	41.1%	38.5%	44.6%
2	42.3%	41.6%	46.9%	45.5%	47.8%	45.6%	44.9%
3	45.5%	44.4%	48.9%	47.8%	50.1%	46.4%	47.2%
4	49.3%	45.9%	47.3%	43.8%	50.0%	47.8%	47.4%
5	41.3%	40.4%	48.0%	50.9%	44.4%	47.8%	45.5%
6	54.6%	48.6%	23.5%	44.9%	46.1%	54.2%	45.3%
7	54.3%	47.8%	49.9%	51.4%	49.4%	50.1%	50.5%
8	47.8%	47.3%	49.2%	46.4%	48.6%	42.6%	47.0%
9	43.0%	52.4%	44.5%	45.1%	46.1%	47.1%	46.4%
10	58.9%	50.6%	46.2%	53.5%	52.5%	55.6%	52.9%
11	50.1%	52.7%	47.8%	49.4%	53.2%	47.4%	50.1%
12	49.6%	48.0%	47.9%	45.5%	52.9%	49.3%	48.9%
Average	49.2%	47.0%	45.6%	47.3%	48.5%	47.7%	

Figure 3.1 Reversal Rates by Participant and Stimulus
Percentage of reversals for each stimulus for each participant. Average reversal rates across all stimuli are in the far-right column, average reversals for each stimulus across all participants re in the bottom row.

3.2 Electrophysiological Results

3.2.1 Reversal Negativity

Given that all visual stimuli were validated as producing the reversal negativity in previous studies, ERPs in response to all three figures were averaged together for all subsequent analyses. Likewise, ERPs corresponding to all three sentences were averaged together after visual inspection of ERP waveforms suggested similar behavior across sentences. Results of the following analyses are shown in Figure 3.2. Mean amplitudes during the time window of the RN were computed for stable and reversal trials for both the visual figures and sentences conditions from a pool of 5 electrodes (26, 42, 43, 53, 54) corresponding to the right-occipital region of interest (ROI) identified for the RN.

In the visual figures condition, a one-tailed dependent-means t-test confirmed that mean amplitudes for reversal trials (M=1.58μV SD=2.65) were more negative than for stable trials (M=2.09μV SD=2.76) in the right occipital ROI, t(11) = -3.58, p < .05. In the sentences condition, a two-tailed test was performed given our lack of a priori hypotheses for effects in this region. Mean amplitudes for reversal trials (M=6.65μV SD=3.03) were found to not significantly differ from stable trials (M=6.93 SD=3.06; t(11) = -1.90, ns) in this right-occipital ROI.

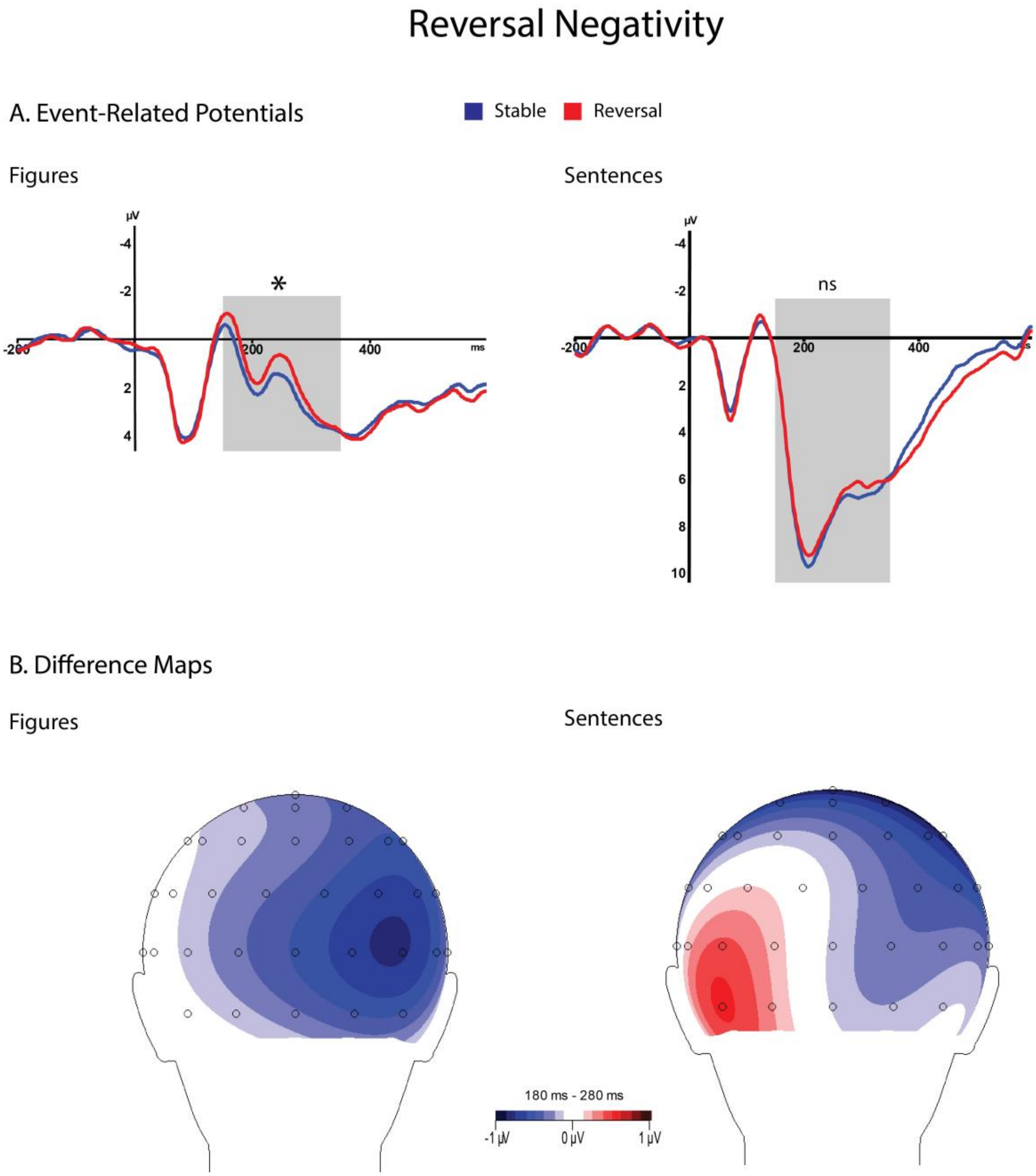


Figure 3.2 Reversal Negativity: ERPs and Difference Maps

(A) Event-Related Potentials obtained in the right-occipital ROI corresponding to the Reversal Negativity in the visual figures and sentences conditions. Mean amplitude from 150-350ms after stimulus onset was compared for reversal versus stable trials in each condition. Stars denote a statistically significant difference in that time window, $p < 0.05$. **(B)** Difference maps showing the mean amplitude difference between reversal and stable trials averaged across a representative time window from 180-280ms.

3.2.2 Frontal Sentence Condition Effect

Results of the following analyses are shown in Figure 3.3. To investigate the negative-going difference observed in frontal regions in the sentences condition, mean amplitudes in the same time window as the visual RN were computed for a left-frontal (7, 18, 19, 33, 34) and right-frontal (3, 9, 10, 21, 22) pool of electrodes and submitted to a two-way repeated measures ANOVA with percept (stable/reversal) and hemisphere (left/right) as within-subjects factors. This analysis revealed a main effect of trial-percept (F(1,11) = 14.75, p < .05). There was no main effect of hemisphere (F(1,11) = 1.09, ns) or interaction effect between percept and hemisphere (F(1,11) = .20, ns).

Given that there was no main effect of hemisphere, data were collapsed across hemisphere and a one-tailed dependent-means t-test confirmed that the mean amplitude for reversal trials (M=1.64µV SD=1.97) was significantly more negative than mean amplitude for stable trials (M=2.55µV SD=1.71; t(11) = -2.90, p < .05). To confirm that this effect was unique to the sentences condition, the same analysis in the figures condition in the frontal ROI revealed no significant difference in mean amplitude between reversal (M=4.76µV SD=1.56) and stable trials (M=4.89µV SD=1.33, t(11) = -.81, ns).

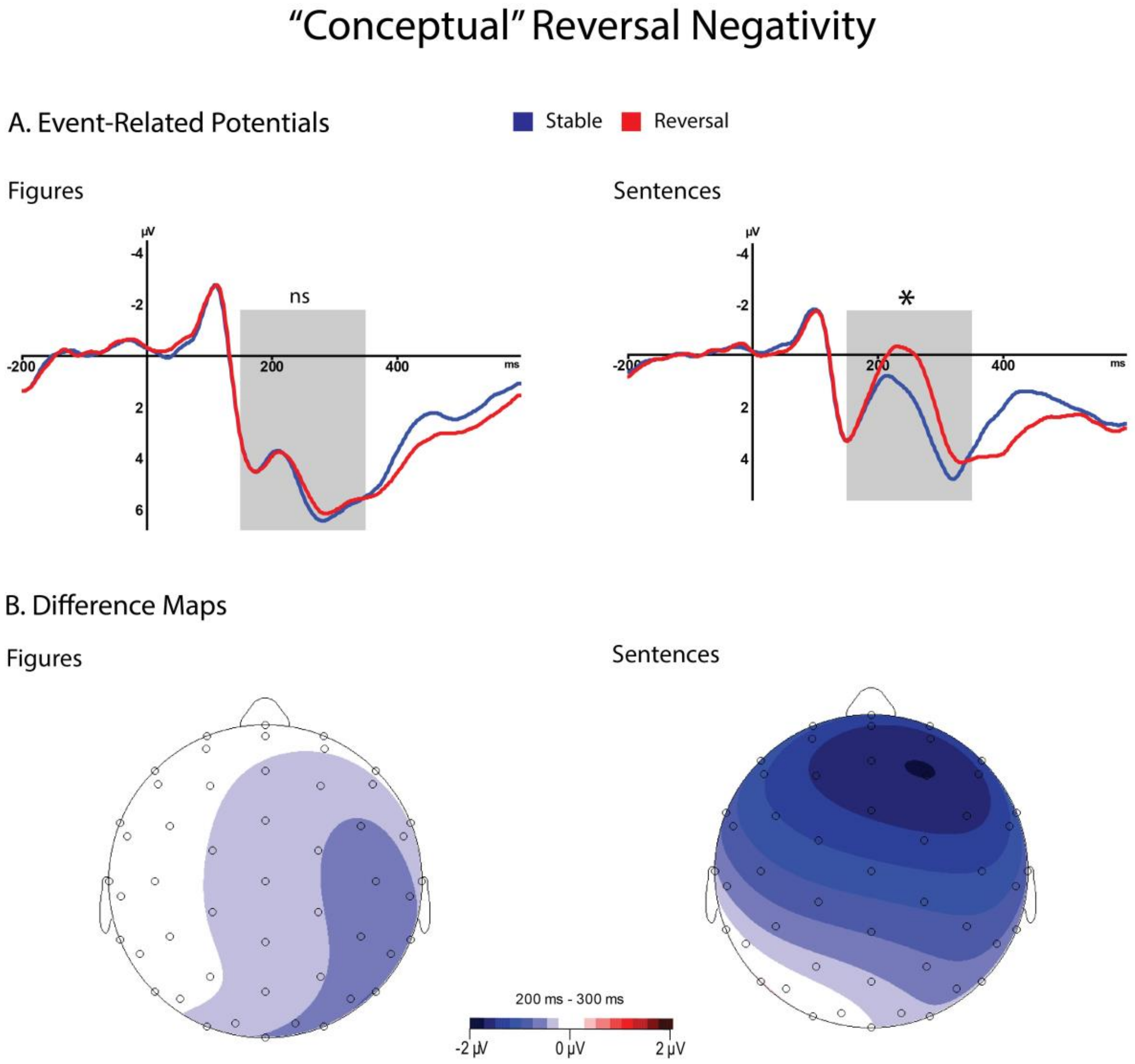


Figure 3.3 "Conceptual" Reversal Negativity: ERPs and Difference Maps
(A) Event-Related Potentials obtained in the frontal ROI corresponding to the "Conceptual" Reversal Negativity in the visual figures and sentences conditions. Mean amplitude from 150-350ms after stimulus onset was compared for reversal versus stable trials in each condition. Stars denote a statistically significant difference in that time window, $p < 0.05$ **(B)** Difference maps showing the mean amplitude difference between reversal and stable trials averaged across a representative time window from 200-300ms.

3.2.3 Late-Positive Component

The Late-Positive Component was investigated using the mean amplitudes from 350-600ms from a pool of seven electrodes (1-7) in each condition corresponding to a central region of interest. Results of the following analyses are shown in Figures 3.4. In the figures condition, a one-tailed dependent means t-test confirmed that mean amplitudes for reversal trials (M=4.81μV SD=2.34) were significantly more positive than mean amplitudes for stable trials (M=4.11μV SD=2.41; t(11) = 3.24, $p < .05$). In the sentences condition, mean amplitudes for reversal trials (M=5.07μV SD=2.34) were again more positive than those for stable trials (M=3.82μV SD=1.74; t(11) = 5.00, $p < .05$).

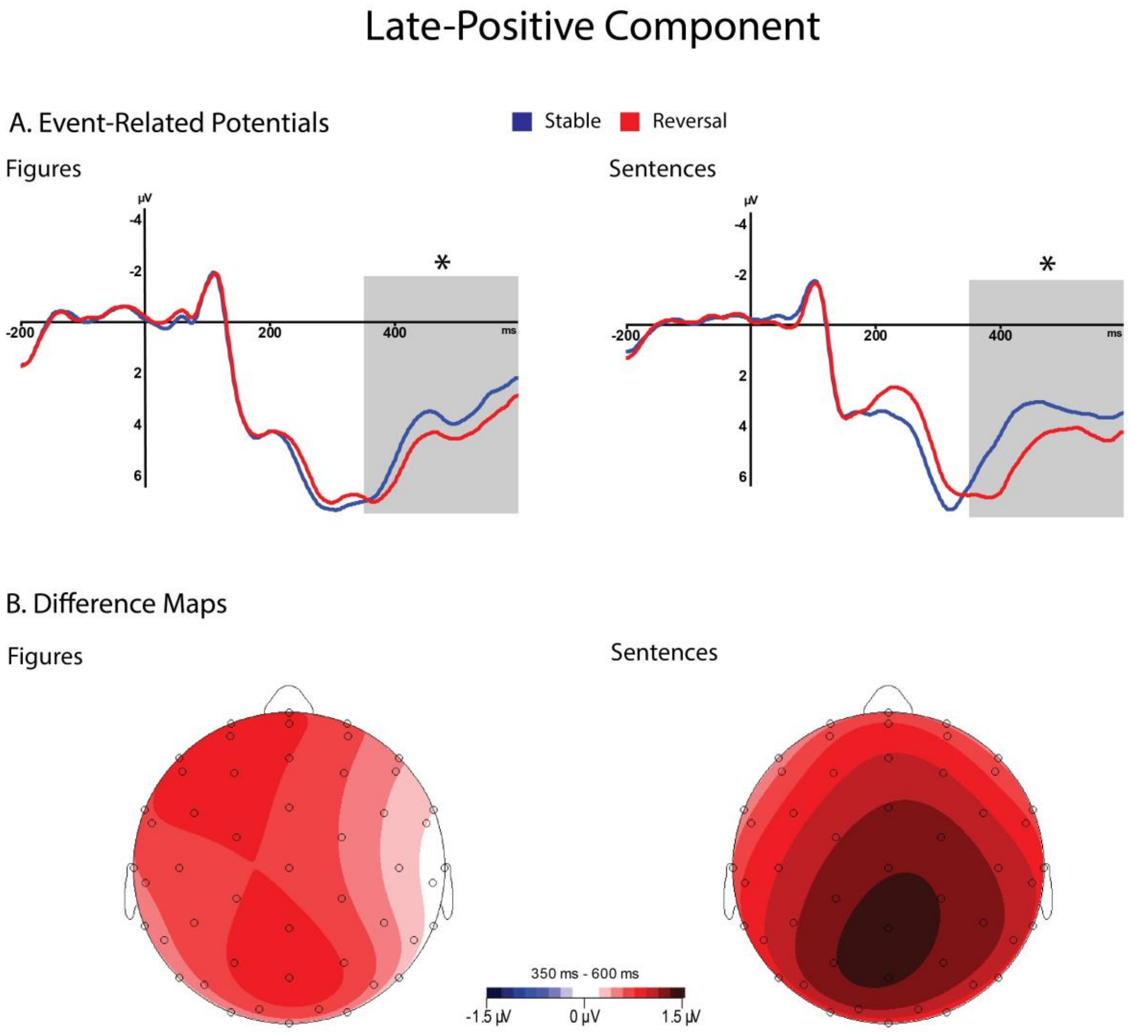


Figure 3.4 Late-Positive Component: ERPs and Difference Maps

(A) Event-Related Potentials obtained in the central ROI corresponding to the Late-Positive Component in the visual figures and sentences conditions. Mean amplitude from 350-600ms after stimulus onset was compared for reversal versus stable trials in each condition. Stars denote a statistically significant difference in that time window, $p < 0.05$ **(B)** Difference maps showing the mean amplitude difference between reversal and stable trials averaged across the entire time window.

Chapter 4: Discussion

4.1 Summary of Results

In the present study, we investigated the electrophysiological signatures associated with the processing of bistable visual figures and linguistic analogues of these stimuli in the form of ambiguous sentences with two valid interpretations. In the visual figures condition, we successfully replicated both the reversal negativity (RN) and late-positive component (LPC) effects using a novel presentation paradigm. Importantly we believe this successful replication validates the use of this paradigm in our extension to the sentence-picture stimuli.

In the sentences condition, we identified a small difference for reversal versus stable trials in the time window of the reversal negativity in the right-occipital region-of-interest, but this effect failed to reach significance (p=.084) in a two-tailed test. Interestingly, however, we observed a large and widespread negativity across fronto-central electrode sites which occurred in a similar time window (~150-350ms) with a similar peak (~250ms) as the visual RN. Additionally, a similar Late-Positive Component was identified in the sentences condition.

4.2 Replication of the Reversal Negativity

As expected, our novel presentation paradigm of an ambiguous figure followed by an unambiguous figure successfully produced a reversal negativity. Previous studies have utilized all-ambiguous, all-unambiguous, or unambiguous-then-ambiguous presentation schemes, and all have produced a reversal negativity. Thus, it is not entirely surprising that our ambiguous-then-unambiguous presentation paradigm also produced a reversal negativity, adding evidence for the robustness of the Reversal Negativity effect.

The lack of an identical Reversal Negativity in the sentences condition is also not surprising, given that neither the sentences themselves, nor the disambiguating images, rely on physical stimulus details to produce bistability or near bistability in the same way as the ambiguous or disambiguated visual figures. It makes sense that sentences paired with drawings and the interaction between these two stimuli would be represented and processed by different mechanisms in the brain than visually ambiguous figures, and as such it would have been improbable to observe an identical effect to the RN in the sentences condition. The question at hand is whether our bistable sentences produce a pattern of activity that suggests that the two interpretations of the sentences are being processed and alternated-between in a manner similar to the perception of bistable figures. As such, rather than looking specifically for a visual RN, we are looking for an RN-like process, and we believe we may have identified such a process.

4.3 A “Conceptual” Reversal Negativity (?)

We believe that the ambiguous sentence stimuli behave in a bistable fashion and that the effect we observed in fronto-central electrodes in the sentences condition may reflect similar processes as the visual RN. We tentatively refer to this effect as the “Conceptual” Reversal Negativity for the remainder of this document. Our conceptual RN shows a similar initial negative deflection at ~150ms compared to our visual RN’s initial deflection at ~120ms, a similar peak latency (~270ms vs ~250ms for the visual RN), and a virtually identical onset of the LPC at ~350ms. Because the sentences condition included a mix of sentence stimuli disambiguated with pictures, it would be inaccurate to refer to this phenomenon as a strictly “linguistic” reversal negativity. Essentially, we believe participants on each trial are forming some type of “conceptual” representation of the sentence which could then be compared to the information conveyed by the disambiguating image.

4.3.1 The Lack of Common Language-related ERPs

The absence of common language-related ERPs such as the N400 and P600 in the present study suggests that rather than relying on the elements of language indexed by those components (i.e. semantic relatedness and syntactic processing, respectively) some other type of relationship between the bistable sentences and picture stimuli is driving the observed effect.

The N400 and P600 are often investigated in active reading tasks such as word-by-word presentations of sentences in which a critical word in the sentence violates semantic or syntactic expectancies. In the previously mentioned picture-sentence verification tasks, a picture is used to establish a context against which a sentence is compared, and violations of the established context occur at the level of a single word in the word-by-word serial presentation of a sentence. In sharp contrast to this methodology, the sentences in our experiment were presented first in their entirety with ample time to form a single stable interpretation prior to the onset of the disambiguating picture. Effectively, all linguistic processing was likely completed prior to the disambiguating image, and it was the meaning of the sentence processed in its entirety that established the representation against which the picture would be compared. A coherent mental representation of an event, delivered via language and compared against an image, is starkly different from comparing the context established by an image to a representation of a sentence which is actively being constructed word-by-word in order to be compared to said context. In contrast, in our paradigm, the semantic and syntactic structure of the sentences are established and stable prior to presentation of the disambiguating image, and the images are presumably being compared to the entire conceptual representation created by the sentence rather than individual unresolved elements of meaning or structure. Because our paradigm lacks this element of on-line language processing at the time of the disambiguating image, our lack of N400 and P600 effects is not surprising when compared against paradigms relying on active sentence processing.

The N400 is also investigated in paradigms where pairs of related or unrelated words are presented in isolation outside the context of a sentence, and the observation of an N400 in these paradigms has been interpreted as evidence of automatic spreading of semantic activations in a lexical network in response to word stimuli. The present experiment clearly differs from these paradigms as well, as it is unlikely that our complete sentences could be interpreted and processed as single "words" or that they would activate semantic networks in the same manner as isolated word stimuli. Although individual words of our sentences and elements of our pictures carry semantic associations critical to the experimental task, we can again make the argument that the disambiguating picture stimuli were not being compared to any individual words of the sentences in isolation, but rather to a holistic representation of a sentence that has finished undergoing processing, and that these picture stimuli themselves are holistic conceptual representations.

Further evidence that our paradigm assessed different properties than those indexed by the N400 in any paradigm is that the negative deflection of our conceptual RN occurred around ~150ms, far before the earliest onsets reported for the N400. Even in paradigms using highly repeated words, which can cause an earlier N400 peak latency, the onset of the N400 itself only begins as early as 200-230ms (Renault et al 2012). Thus, although each element of our disambiguating images (such as the chicken as food or the live chicken) has its own semantic associations, which in theory could be compared to the established semantic representations of the unambiguous sentence (such as that critical word "chicken," which is the source of the ambiguity of the sentence), it does not appear that these relationships of semantics at the word level, at least those indexed by the N400, are what is driving our effect. Rather, we believe that some other level of representation and meaning is responsible.

4.3.2 The N300: A Possible Explanation?

In attempting to interpret our results, we identified another language-related ERP component in the literature which is of potential relevance: the N300. Similar to the N400, the N300 is observed in response to semantic incongruencies and is sensitive to word-relatedness, but the N300 is specific to visual stimuli and may reflect recognition and categorization of visual objects and matching of these objects with stored semantic knowledge (Maguire 2012, Hamm 2002). Time windows for analysis of the N300 are variable, beginning as early as 125ms (Maguire 2012) and ending as late as 400ms (Debruille 2012), with peaks reported between ~250ms and ~300ms in these studies. Also of relevance is that the N300 is normally accompanied by a positive-going difference at occipito-temporal or parietal scalp sites, with the neural generators of the N300 potentially being localized to these regions (Debruille 2012, Maguire 2012).

McPherson and Holcomb (1999) investigated N300 and N400 effects in the context of pairs of related (“burger” and “fries”) or unrelated (“cat” and “chair”) photos of real objects. They observed an N300 in the 225-325ms time-window, as well as a subsequent N400 effect in response to incongruous image pairs. However, the authors do not report on any posterior positivity effects, and visual inspection of the presented waveforms suggests that they did not observe an LPC. Barrett and Rugg’s (1990) study using related and unrelated pairs of line drawings, upon which McPherson and Holcomb’s study was based, report similar results. Hamm et al (2002) presented word primes of a category (such as “dogs”) followed by images of either a member of that category (“Labrador”) or a member of a different category (“pigeon”) and recorded ERPs time-locked to the image onset. They observed an N300 effect in the time window from 244-288ms, corresponding to the component’s peak, when comparing category mismatches (“dog” followed by an image of a pigeon) versus matches (“dog” followed by an image of a Labrador), and also report a subsequent N400 following the N300. An accompanying posterior positivity to the N300 appears to be visible in the data across anterior-parietal, anterior-temporal, and occipital electrodes. Again, no LPC effect appears in the waveforms.

Similar to Hamm et al.’s procedure, Mazzerole et al (2007) presented word primes for a category (fruit, tool, etc) or specific category member (orange, hammer) followed by a picture that either matched or mismatched with the word prime. The authors again observed an N300 effect for mismatch trials using an analysis window from 250-350ms. In contrast to Hamm et al.’s study, no subsequent N400 was reported. Additionally, the authors in this study report an LPC effect between 400-600ms, however, the amplitude of LPC they observed was not statistically significantly sensitive to the congruence of the word prime and picture. Instead, their LPC was related to object category, being larger for natural objects than artificial objects. The authors do report that despite the lack of a significant effect of congruence on LPC amplitude, congruent trials produced an earlier LPC peak latency than incongruent trials. It is unclear how to interpret these LPC findings given the lack of a statistically significant effect of congruency on LPC amplitude within the specified time window, despite visual inspection of ERP waveforms suggesting that such an effect may be present.

Most similar to the present experiment, Maguire et al (2012) conducted a study in which a word prime specifying either an object ("sandwich") or an action ("eat") was followed by a color drawing of a scene which depicted both an object and action that could be either congruent (a woman eating a sandwich) or incongruent (a boy opening a door) with the context established by the presented word prime. Similar to our experiment, they found a negative-going difference for incongruent versus congruent stimulus pairs across frontal electrode sites in a time window from 125-300ms, roughly similar to our analysis window of 150-350ms. In this same time window, they found a corresponding parietal positivity for incongruent trials as compared to congruent trials. Like Barret and Rugg, McPherson and Holcomb, and Hamm et al, the authors also observed an ongoing negativity for incongruent trials into the N400 time-window (300-500ms), and the data do not appear to show an LPC. Of relevance to our experiment is the finding of an N300 for verb congruency mismatches, as verbs could be considered more similar to complete events as conveyed by our ambiguous sentences. This verb congruency finding suggests the N300 may be sensitive to other aspects of meaning aside from simple object recognition, and also of interest is that visual inspection of the authors' ERP waveforms suggests that these verb congruency N300s had a later initial negative deflection around 200ms compared to their noun congruency N300 effects which began at approximately 125ms.

Comparing our data with the characteristics of the N300, the effect we observed between 150-350ms clearly lies within the range of N300 time windows reported in the literature (125-400ms). The peak of our effect at ~270ms is also consistent with a peak between 250-300ms, and our effect shares a similar fronto-central scalp distribution. A potential divergence from the typical N300 findings is in the posterior positivity that accompanies the N300. In the present experiment, we observed a weak positivity in the same time window as the N300 in left-occipital/posterior-temporal electrodes which may correspond to the aforementioned positivity effects, although it is difficult to draw any clear conclusions because this comparison is based only on visual inspection of the subsets of data presented by these authors and visual inspection of our own data.

An additional contrast between our results and the aforementioned studies of the N300 is in our finding of an LPC. Neither McGuire et al., Hamm et al., Barrett and

Rugg, nor McPherson and Holcomb report or appear to show an LPC effect. Mazzerole et al. do report an LPC, which is of interest considering the similarity of their paradigm to that of Hamm et al where no LPC appears. A key difference between Mazzerole and Hamm's highly similar procedures is that participants in the Mazzerole study were required to indicate after each word-picture pair whether the pair matched or mismatched via button press, whereas in Hamm's study, participants passively viewed the word-picture pairs without providing a response. The requirement for a participant response being the cause of the observed LPC in Mazzerole's study could potentially offer an explanation as to why we observe an LPC in our own study, as well as potentially to why Mazzerole et al. do not report finding an N400. This explanation, however, is insufficient, because Barrett and Rugg, McPherson and Holcomb, and Maguire et al. also required participant responses in their studies, yet still reported subsequent N400s following the N300 and did not observe LPC effects. Given these results it is unlikely that participant response alone is sufficient to explain the presence or absence of an LPC.

The LPC effect observed in our experiment may be directly due to task demands, an interpretation which fits with the LPC's functional interpretation as being related to the updating of the contents of working memory. Because our paradigm relies on trial-by-trial comparisons of the same ambiguous and unambiguous stimuli, it makes sense that encoding of each stimulus into working memory, as indexed by the LPC, would occur. In contrast, the N300 studies discussed above all used a large number of stimuli such that pairs of related and unrelated pictures or word-picture pairs were novel (i.e. not repeated) over the course of the experiment, thus removing any specific demands for storage of the particular object representation into memory. A possible exception is Mazzerole et al.'s study, which used a large number of stimuli, but presented word-picture pairs in blocks of ten trials, with each block consisting of five match and five mismatch trials, rather than presenting stimuli completely randomly as in other studies. Although trials within blocks were randomized, it is possible that participants became sensitive to the equal number of matches and mismatches per block which may have produced an element of expectancy and predictability as to whether subsequent trials were more likely to contain matches or mismatches. Although speculative, this predictability could have led participants to actively store information regarding the congruency of the trials within blocks, which could produce an LPC. This speculation must further be qualified by the fact that the LPC observed in Mazzerole et al.'s study was not significantly modulated by congruency, although congruency effects do appear to be potentially present in their ERP waveforms despite failing to reach statistical significance.

The same logic regarding the absence of an N400 given the differences in our paradigm from typical N400 paradigms largely applies to these studies of the N300 as well. Rather than presenting word or image primes representing a single concept (such as "dog" or "eat"), we presented full sentences against which the disambiguating image would be compared, which in theory produce a more complex conceptual representation. An alternative interpretation of our results in terms of simpler types of semantic relationships underlying the N300, is however, largely possible.

In the sentence "The chicken is ready to eat," the critical ambiguous word is "chicken" which can be assigned either the role of the agent of the verb "to eat" or as the object of the same verb. Despite "chicken" being one word, it could be that the word has two stable semantic or representational interpretations in the brain (as an animal or as food), which in essence is why we believed this sentence would act as a linguistic analogue to bistable visual figures. In "I saw her duck," the word "duck" can likewise be interpreted as either a verb referring to the action the woman is performing or as a noun signifying the animal belonging to the woman. Similar to "chicken," the word "duck" may be stably and reversibly interpreted as either a noun or verb. It is plausible then that the observed effects resulted from semantic incongruencies of these critical words with the presented images, despite the fact that these incongruencies are the result of a single word having multiple meanings, rather than being an outright mismatch of meaning between two different words, which is an interesting result in itself.

Also worthy of consideration is the fact that the disambiguating chickens in the images were embedded in the context of an entire scene with other elements, such as steam and a fork-wielding hand, or grain and a benevolent snack-offering hand. The disambiguation provided by the duck images relied more heavily on the context of the entire image, depicting either a woman performing an action or a woman's relationship to an animal. In both of these cases it is unclear how much of a difference the context of the images may have impacted the observed results, however. In the chicken images, the disambiguating information was located in roughly the same spatial location, and that attention could be directed only to this one region in order to make a comparison to one's interpretation of the sentence, theoretically without active processing of the rest of the image. The duck images, despite depicting more complex relationships to establish meaning, were starkly different in their composition such that these basic visual differences may have been sufficient to determine whether the image matched or mismatched with one's interpretation of the sentence. Despite these differences, these sentences produced similar reversal effects, and the issue of the contribution of processing of the scene in its entirety versus the processing only of critical sub-elements and features remains open.

This interpretation of the results thus far may fit more parsimoniously within the framework of our observed effect being a variant of the N300, rather than a wholly novel effect, but the fact that the sentence "She hit the man with the umbrella," also produced the observed frontal reversal effect provides evidence against the idea that the effect is due only to the mismatch in meaning of one key sentence and picture element. Unlike the previous two sentences which rely on the semantic ambiguity of the words "chicken" and "duck", "She hit the man with the umbrella" relies on a syntactic ambiguity in the attachment of the phrase "with the umbrella." In this sentence there is no single word with an alternate interpretation that can disambiguate the meaning when presented visually. Instead the entire phrase "with the umbrella" and its attachment to either the verb "hit," modifying the manner of the action, or to the noun "man," as a modifier of the noun, must be processed in order to resolve the ambiguity. The disambiguating drawings paired with this sentence reflect the greater subtlety of this ambiguity, and the visual content of these drawings is nearly identical aside from the placement of the umbrella. In order to make a comparison between one's interpretation of the sentence and the presented image, it is necessary to process not only the object meanings in the sentence and image, but also their relationships in order to form a complete representation, which is likely a qualitatively different computation than extraction of meaning. Our lack of a P600 effect also suggests that the computation being performed here is not specifically related to the syntactic structure of the sentence per se. This suggests the possibility that rather than being caused by sensitivity to the meanings of specific subcomponents of the presented sentences and images, the observed effect may reflect sensitivity to a broader level of meaning that can encompass the conceptual extent of an entire sentence and visual scene.

4.3.3 Summing Up

Ultimately, these arguments alone are not sufficient to conclude that our observed effect is indeed novel, and it may instead be that our results show that the N300 is sensitive to broader scales of meaning than those used in the paradigms which have reported it thus far. Our lack of an N400 effect, however, does still suggest that something different is being assessed in our experiment as compared to these N300 studies, although the lack of consensus and difficulty of interpreting the presence and occasional absence of N400 effects in the identified N300 literature makes it unwise to draw any firm conclusions.

To further frame the discussion of our effect's potential similarity to the N300, it is necessary to consider the divergence in our methodology from the typical N300 paradigms, where semantic relatedness is assessed among many word-picture or picture-picture pairs. In contrast to these paradigms, we focused on sentences with "bistable" interpretations, and we presented individual sentences repeatedly over the course of an entire block in order to investigate "reversals" in the interpretations of these bistable sentences in a manner similar to the reversal of bistable figures. The fact that the observed effects and their time course in our paradigm matches so closely with those observed effects for the visual RN suggests that similar processes as those reflected by the RN are at play.

The crux of the issue is whether the effect we observed reflects a shift in the configuration of the conceptual representation of a sentence, in the same way that the visual RN is thought to reflect a shift in the perceptual configuration of a visual figure, or whether our effect is simply an N300 reflecting some incongruence of meaning between two presented stimuli in the context of the given task. It may be the case that our effect is an N300 and that the N300 can index the same type of switching between interpretations of bistable sentences in a manner similar to the visual RN, in which case the N300 is the analogue of the visual RN in the context of sentence-picture stimulus pairs. It may also be the case that, under the constraints of our experimental paradigm, our observed effect is an N300 which, instead of reflecting any shift in a conceptual representation of a bistable sentence, is simply a reflection of the detection of some type of incongruence between the current mental representation of the sentence and the disambiguating image, whether that be at the semantic level or some broader conceptual level. If this is case, the N300 could not be considered an analogue of the visual RN, and we are left with the possibilities that either our task failed to sufficiently parallel the visual reversal negativity paradigm, or that representations of bistable sentences simply behave differently in the brain than bistable figures, and that any similarity in our results and typical RN results is due to task constraints.

Distinguishing between these possibilities and reaching a firm conclusion about the relationship of our observed effect to the N300 is impossible without further research, but we have several reasons to believe that our observed effect is in fact similar to the visual RN and that it is sensitive to changes in the current conceptual representation of these bistable sentences. First, our new presentation paradigm produced a typical visual RN in the visual figures condition, and we employed the exact same paradigm in the sentences condition. If the effect we observed in the sentences condition is due to task demands and simple comparisons of congruency between stimuli, this may have implications for the role of task demands in producing the visual RN effect, a possibility discussed below. Second, the consistency of our observed effect across three different sentences relying on different types of ambiguity, whose disambiguating images likewise differed in terms of visual similarity, suggests that our effect is sensitive to a wider scope of meaning than simple semantic relatedness, as has been typically explored in N300 studies. Finally, there are important potential differences between our observed effect and the N300 effects observed in the five aforementioned studies: first, we did not observe the subsequent N400 that four studies reported. Second, our effect potentially lacks the accompanying posterior positivity reported by two studies, and third, our effect seemingly has an earlier onset than those observed in the aforementioned studies, aside from McGuire et al.'s onset of ~125ms for object incongruency, which must be contrasted with the apparent onset at ~200ms for verb incongruency. Taken together, these results tentatively suggest that our observed effect may be distinct from the N300 effect.

4.4 One Effect or Two?

The marginally significant negativity in the sentence condition that we observed in the same right-occipital ROI as the visual RN is interesting and contrary to the expectation that reversals in our sentence-picture paradigm would rely on largely different brain processes and representational mechanisms than those involved in the visual RN. Individual participant data suggests that there may be a difference in how and where the difference between reversals and stable trials in the sentence condition manifests. The majority of participants (9 of 12) demonstrated the large frontal negativity effect between 150-350ms with no apparent differences in occipital electrode sites corresponding to the ROI for the visual RN. Three participants, however, also showed a negativity effect in right-occipital sites resembling that of the RN for visual figures. Of these three participants, two also showed the frontal effect in addition to this back-of-the-head effect, and one failed to show any frontal effect.

Further data collection is necessary to determine whether this back-of-the-head effect is actually significant, and whether additional participants show similar effects and/or a lack of the frontal effect. It is possible that these differences are due only to random inter-individual variability unrelated to the underlying cognitive mechanisms at work in the task, or to the widespread nature and large magnitude of the frontal effect making a negativity detectable even at posterior scalp regions. Of interest here is that we observe a left-lateralized positive difference for reversal versus stable trials in the back of the head in the same time window of the frontal effect and the marginal right-occipital effect. Preliminary inspection of scalp maps suggests that this left-lateralized posterior positivity may be the opposite end of the dipole that is generating the slightly right-lateralized frontal effect. One could hypothesize that spreading of the positivity generated in this left-occipital region might be masking any negativity that could be occurring for reversals in the right occipital ROI, which is an intriguing prospect if it were the case that reversals mediated by language and disambiguated by dissimilar visual stimuli are causing a similar pattern of occipital activation as those for highly-visually-similar bistable figures. Speculation based on a few deviant data points in our relatively small sample is ultimately of limited informative value, but it does indicate that there may be further interesting aspects of our conceptual reversal effect that may become apparent as a clearer picture develops with further data collection.

4.5 Task Strategies and Methods of Representation

A point of interest in assessing between-participant variability in the sentence condition is that participants may have employed different types of strategies when performing the task. Participants were only instructed to “form a stable interpretation of the sentence” on each trial, with no explicit instructions as to how that interpretation ought to be formed. Some participants noted to the experimenter that they were visualizing a mental picture corresponding to one of the two disambiguating drawings on each trial, while other participants noted more language-based strategies such as thinking “chicky chicky” or “steamy steamy” or “cute/gross chicken” on each trial. Additionally, some participants indicated that they conceptualized the task as trying to guess which stimulus would appear next, while others indicated that they were relatively passive and allowed reversals to happen naturally with no goal of predicting the next stimulus. No data was formally collected to assess strategies, and as such, a systematic investigation of the effects of task strategy on individual results is not possible at this time. Despite this variablity in task strategies, we observed the frontal reversal effect in the sentences condition in all but one participant. Nevertheless, the variability in task strategy raises several interesting issues that could be relevant in further data collection, or as investigations worth pursuing on their own.

4.5.1 The Role of Language

Given that the experimental task in the sentences condition consisted of both written sentences and visual images, it is unclear how much of the observed effect is due to language processing per se, versus image processing, versus some blending of the two into a general “conceptual” task. Because only one ambiguous sentence was used for each block of 180 trials, it is possible that there was an over-learning effect in which participants stopped effortfully reading and carefully processing the sentences, and instead opted to select one or the other possible interpretations with minimal, if any, language processing. Participants’ reports support this notion, with some participants indicating to the experimenter that they felt they were no longer actively reading the sentences, while other participants indicated that they were attempting to read the sentences consciously on each trial for the duration of the experiment. No formal assessment was made of thoroughness of reading, and as such our dataset likely includes a mix of both effortful and minimal reading approaches. As is the case for the overall question of the relevance of task strategies, the frontal reversal effect was observed in all but one participant, despite this inter-individual variability.

It is, however, implausible that language processing played no role whatsoever in the current results, even in participants who reported not processing the sentences in an effortful manner or who adopted primarily mental visualization strategies. Orthographic processing of word forms can occur even in the absence of awareness, and semantic properties as indexed by the N400 appear to be accessed for task-irrelevant words even while performing a demanding distractor task under conditions of high visual load, as long as subjects are simply aware of their presence (Schelonka et al 2017). In the current study, the sentences were presented for a relatively long duration of 800ms and were the only stimulus on the screen, so it is implausible that there was no processing of the sentence whatsoever, even in participants who were paying a minimal amount of attention to the sentences, given the automaticity of language processing.

4.5.2 A Possible Role of Mental Imagery?

Individuals’ capacity for generating mental imagery is variable, and scales such as the Vividness of Visual Imagery Questionnaire (VVIQ) attempt to assess the variation in this purely subjective phenomenon (Marks 1973) . Brain imaging studies using fMRI comparing mental imagery versus actual perception have found substantial overlap in activition in frontal, temporal, and anterior parietal regions, with differences between the two tasks arising primarily in posterior parietal and occipital regions (Ganis et al. 2004).

Additionally, activity in the visual cortex has been found to correlate with individuals' vividness of imagery as assessed by the VVIQ (Cui et al. 2006). Recent research indicates that participants with low-vividness of imagery as assessed by the VVIQ may activate a more diffuse, widespread network of brain regions than those with high-vividness, whose activations are more concentrated, and that many areas activated only in the low-vividness group display a negative relationship with vividness of imagery (Fulford et al. 2017).

ERPs are not particularly well-suited to studying mental imagery due to the difficulty of time-locking to the fully internal and temporally extended process of generating a mental image, although some literature does exist on the topic. Farah et al. (1989) recorded EEG time-locked to the onset of a visually presented word in two conditions: one in which participants were instructed to silently read the word and another in which they were instructed to generate a mental image of the word's referent. As compared to the passive reading condition, in the imagery condition the authors observed a late, slow positivity from ~600-1000ms, maximal at occipital and posterior temporal electrode sites and lateralized to the left side of the scalp. Shen et al. (2015) found a negative-going frontal effect of imagery between 200-750ms with word-by-word presentations of literal versus abstract sentences, with the amplitude of the effect being correlated with participants' VVIQ scores. Additionally, high-VVIQ participants displayed a similar imagery effect when reading sentences with unfamiliar or familiar metaphors as compared to literal sentences, with the effect being more temporally extended for unfamiliar metaphors, whereas no imagery effect was present for low-VVIQ participants in either metaphor condition, although low-VVIQ participants appeared to show a posterior N400 effect for unfamiliar metaphors. The authors interpret this frontal effect as suggesting that participants with high imagery abilities may recruit frontal sensory-motor areas when visualizing, although this interpretation should be considered cautiously, given that an ERP effect's location on the scalp does not necessarily imply that the neural generators responsible for that effect are located in a similar region of the brain (See Luck 2014 for discussion).

These results indicate that mental imagery has observable ERP effects and that these effects may be sensitive to individual variations in vividness of imagery. Thus it is possible that imagery differences may modulate some aspects of the observed effects in the present experiment, although again, no formal assessments of vividness of imagery were made. As a point of interest, the single participant who failed to show the frontal conceptual RN effect reported that they lack any distinct capacity for mental imagery. As a cautionary counterpoint, one participant who reported having extremely vivid mental imagery and noted that they were using a visualizing strategy in the task showed a typical pattern of results with no distinguishing features. Much like the role of language, it is difficult to determine what explicit role imagery might play given the combined linguistic and visual nature of the task. Additionally, it is difficult to hypothesize what effect a greater vividness of imagery should correlate with in our paradigm, given both the novelty of our paradigm and the limited previous ERP research involving mental imagery. Nevertheless, the potential role of imagery in the formation of representations of sentences in a task such as this is an interesting question for future research.

4.5.3 The Role of Expectancy (and a Brief Clarification on Reversal Rates)

In contrast to most studies of the reversal negativity where participants passively view serial presentations of an ambiguous stimulus and are asked to indicate when they perceive a reversal, participants in the present experiment were prompted to respond to individual pairs of stimuli and were instructed to attempt to perceive each interpretation of the ambiguous stimuli roughly half of the time, if possible, with the goal of obtaining roughly an equal number of stable and reversal trials for each interpretation of the ambiguous figure. Average reversal rates in the visual figures and sentences conditions were 47.1% and 47.8% respectively, indicating that participants reliably had more stable trials than would be expected by random chance. These values are still reasonably close to 50%, however, and the lowest reversal rate recorded for any participant was 44.5%, which does not suggest that substantial guessing of the coming stimulus was possible to a degree that would be problematic. Because each disambiguating stimulus was presented an equal number of times per block (90 times per block of 180 trials) and randomized without replacement, this slightly above-chance performance is not surprising, and it is plausible that participants could anticipate whether one disambiguating image was more likely to appear than the other on a trial-to-trial basis, especially after series of trials where a certain stimulus happened to be repeated several times in a row or near the end of blocks. This result fits with some participants' reports that they were attempting to guess which stimulus would come next, and suggests that other participants may also have been influenced by these probabilities, even if only subconsciously. This raises the issue that some of our observed effects, rather than being strictly due to reversals in the interpretations of the ambiguous stimuli, may have been driven by guessing or expectancy. This issue of violation of expectancy contributing to the reversal negativity effect has been explored before, however, and we do not believe that expectancy is a critical determinant of our results.

If violations of expectancy were driving the RN effect, one would predict that participants with less-frequent reversals would show a greater RN amplitude given the greater saliency of these less-frequent reversals, as expectancy-related ERPs are found to be greater in amplitude as the frequency of a deviant stimulus decreases. Davidson and Pitts (2014) explored this possibility in regards to their finding of an auditory analogue of the RN (the aRN) and found no significant differences in aRN amplitude for participants with the lowest reversal rates compared to the highest rates. The finding that this auditory RN is not dependent on violation of expectancy provides evidence that the aRN is not a variant of the auditory Mismatch Negativity (MMN) effect, which occurs at a similar latency and scalp distribution and which is sensitive to frequency of deviant tones. One may question the relevance of an auditory reversal finding to our current reversals, but given that the auditory system demonstrates sensitivity to violations of expectancy via the MMN and that the aRN appears to be distinct from these effects, this suggests that reversal effects in other domains may likewise be independent from any effects related to violations of expectancy.

Jimenez's 2017 thesis study of the reversal negativity using the Rat-Man figure conducted an even more direct investigation of the effects of expectancy on reversals. In the study, disambiguated Rat-Man figures were presented in two conditions, one in which the disambiguated variants were fully randomized and thus unpredictable, and another in which the stimuli were presented in a consistent pattern, making them fully predictable (i.e. Rat-Rat-Man-Man...). An RN and LPC were found in both conditions, even though all stimuli and reversals in the predictable condition would have been fully expected and anticipated by the participants. Together these results make a strong argument that expectancy is not the driver of the RN.

Our reversal rate of 47.5% is actually substantially higher than typical reversal rates of ~25-35% reported in much of the RN literature, which is due to our paradigm of presenting an ambiguous stimulus first, followed by a disambiguating stimulus, and which we do not believe to be a concern in terms of the validity of our findings. In the case of most studies of the RN which utilize serial presentations of unambiguous stimuli, reversals are purely endogenously driven, and rates of reversals can be influenced by many factors such as inter-stimulus interval and stimulus duration (Kornmeier et al., 2007). In the case of serial presentation of unambiguous stimulus variants, reversal rate is entirely under the control of the experimenter in terms of what sequence of figures is presented. In our experiment, because each unambiguous stimulus variant was presented randomly and an equal number of times per block, our overall reversal rate (the rate irrespective of bias towards one interpretation of the ambiguous figure or another) should in theory always be roughly 50%. The instruction to try to perceive each interpretation roughly half the time was given to obviate the risk of extreme biases causing an imbalance in the number of trials that each possible interpretation of the ambiguous figure contributed to the analysis. Stimuli were pilot tested to confirm that both interpretations were easily perceived and minimally biased, and even in an extreme case such that a participant would perceive the "rat" interpretation of the ambiguous Rat-Man on every trial in a block of 180 trials, this would result in ninety stable trials in response to the unambiguous "rat" and ninety reversals in response to the unambiguous "man" for an overall reversal rate of 50%. Due to randomization, this expected reversal rate of ~50% holds for any given distribution of percepts of the ambiguous stimulus, and our obtained reversal rate of 47.5% bears this out.

When considering the visual figures condition, this higher rate of reversals, occurring roughly 50% of the time due to our experimental design, provides further evidence that expectancy is not the main driver of the RN, as both reversals and stable trials were roughly equally likely. Additionally, the fact that we obtained a typical pattern of both RN and LPC with our novel paradigm, with no other apparent ERP effects, suggests that any guessing strategies or stimulus predictability that may have resulted from our paradigm did not impact the results in this condition. Despite the likely inter-individual variability in the degree to which participants were passively viewing versus anticipating which disambiguating stimulus would appear next, potentially biasing their interpretations of the ambiguous figures, the RN and LPC effects we observed were consistent across participants, again suggesting that any effect of guessing was minor. Because the sentences condition was identical in experimental design to the figures condition, we can tentatively conclude that guessing and expectancy likewise did not play any substantial role in our observed results, and that guessing and violations of expectancy are not the cause of our conceptual RN effect.

4.6. Reconsidering Reversals More Broadly

One might argue that the overall design of these reversal paradigms causes the two potential interpretations of an ambiguous stimulus or its two unambiguous variants to become functionally linked due to task-related constraints. It could be argued that this functional linkage and constrained expectation that one or the other percept will appear is what is causing the Reversal Negativity, instead of the RN being the result of switching between two representations that are intrinsically linked by virtue of ambiguity. It would be easy to apply this interpretation as a criticism against our conceptual RN truly reflecting reversals of interpretations at a conceptual level, but the same argument would have to be applied to visual reversals, and the possibility that the visual RN is entirely unrelated to ambiguity and only due to task constraints seems very unlikely.

If the visual RN were in fact unrelated to ambiguity and were instead due entirely to functional linkages between stimuli created by task constraints, it would theoretically be possible to observe an RN to unambiguous but visually similar stimuli, such as line drawings of a baseball and a basketball. One could present such similar but unambiguous line drawings in a random order and observe if an RN is produced on trials where the stimulus switches from one to another, as has been demonstrated in similar paradigms with unambiguous variants of bistable stimuli. If such an RN is still found, one might argue that these stimuli are still intrinsically linked by both being members of the category “ball”. Thus, to be a fully convincing demonstration, one would have to present two entirely unrelated, visually-dissimilar stimuli, such as a square and a drawing of a plant. If an RN were found here, it would provide strong evidence that task constraints and functional linkages, rather than ambiguity, are the cause of any RN effect, and would effectively undermining the interpretations of the entire body of literature examining the RN thus far. If the visual RN were caused by such task constraints, it would be hard to argue that our conceptual RN reflected anything different. To confirm this, however, similar control experiments such as presenting an irrelevant unambiguous sentence such as “The man walked his dog” and instructing participants to anticipate one of two subsequently presented irrelevant images, such as a child climbing a tree or a woman sitting in a car, would be necessary.

To the best of our knowledge, no one has performed such absurd control experiments, and it would be an interesting result in itself if the functional linkages via task constraints hypothesis turned out to be true. Thankfully, however, we have ample other evidence that the brain is in fact sensitive to ambiguity specifically and that this is not the case. In the context of reversal paradigms, an early reversal positivity (RP) around 130ms is found in response to endogenous reversals of bistable figures and not to reversals of disambiguated variants, which suggests that the brain is in fact sensitive to detection of ambiguity specifically. In regard to the reversal negativity, Kornmeier and Bach (2014) did not find a typical RN peaking around 250ms for reversals of Boring’s Old/Young Woman stimulus. Instead, a negative difference for reversals versus stable trials was found around 170ms, which corresponds to the face-sensitive N170 component. The authors identified an RP in response to this stimulus as well, and interpret these results as suggesting that ambiguity is initially detected by the visual system around 130ms, and that these ambiguities are resolved by early visual and object-specific visual areas during the time period between 130-260ms, with the Old/Young Woman’s ambiguity being resolved earlier in time compared to geometric figures, around 170ms, due to the presence of face-specific regions in the brain which can complete enough of the processing necessary for disambiguation by this time.

If functional linkage and mere expectation of one stimulus or another were driving the RN, one would expect that simple visual discrimination between the two stimuli would be sufficient for determination of whether a reversal occurred, and thus that the latency of signatures of reversal effects according to this explanation would be modulated by ease of stimulus discriminability. The fact that exogenous reversals of disambiguated bistable figures produce earlier RN peak latencies than endogenous reversal of ambiguous figures supports this idea because unambiguous variants are likely to be discriminable earlier, but also supports the RN's typical interpretation because a coherent shift in the current perceptual configuration of the stimulus may be able to occur more quickly as a result of this easier discriminability. The pattern of results we observed in our sentence condition provide valuable evidence in resolving this debate. Due to the relative great visual dissimilarity of our disambiguating drawings compared to bistable figures, particularly in the case of "I saw her duck," discrimination between alternatives should be able to be accomplished more quickly, based on lower-frequency spatial information, and we should thus see sooner RN effects. However, we observe slightly later effects in terms of both onset and peak latency for our conceptual RN in response to ambiguous sentences. We see no clear reason why visually-dissimilar, easily-discriminable drawings should produce later signatures of some task-constrained violation of expectancy as compared to more visually similar and difficult-to-discriminate ambiguous stimuli, and thus conclude that resolution of ambiguity specifically and formation of stable representations of an ambiguous input is a critical factor in both the visual RN and our conceptual RN.

As a final point illustrating the brain's specific sensitivity to ambiguity, Kornmeier et al. (2016) considered three types of ambiguous stimuli: Necker Lattices, which rely on geometric ambiguity, stroboscopic alternative motion (SAM) stimuli which produce reversible patterns of the perceived motion of two dots, and Boring's Old/Young Woman, which the authors describe as a "semantic" ambiguity. Necker lattices can be made more or less ambiguous in a continuous fashion via manipulation of back-layer luminous, as can SAM stimuli by changing the aspect ratio of the dots' alternations, and the authors constructed stepwise variations of these stimuli ranging from fully ambiguous to fully unambiguous. While such a continuous manipulation is not possible with the

Old/Young Woman, ambiguous versus disambiguated variants were still compared. Stimuli were presented in a repeated presentation paradigm as in typical RN studies, but rather than comparing reversal to stable trials, the authors considered only reversal trials and compared ERPs to each stimulus across its levels of disambiguation.

For all three stimuli, the authors identified a centro-frontal positivity at approximately 200ms and a centro-parietal positivity around 400ms whose amplitudes were inversely correlated with degree of ambiguity. That is, the more clearly unambiguous the stimulus was, the greater the amplitude in these two time windows. The authors name these effects the "ERP ambiguity effect" and suggest that, because of their consistency in three very different types of ambiguity, these effects may reflect some mechanism of ambiguity detection at an abstract cognitive level of processing that operates with great generality across different types of ambiguities.

4.7 Conclusion

In the present experiment, we identified an ERP effect in response to reversals of interpretation of bistable sentences which closely mirrored the time course of the established Reversal Negativity for bistable visual figures. We interpret this finding as suggesting that more conceptual types of ambiguities, such as those conveyed by ambiguous sentences, may result in multiple stable representations in the brain that behave in a bistable fashion, similar to the multiple possible representations of bistable visual figures. If our preliminary interpretation of our findings as a "conceptual" reversal negativity is verified, it would further demonstrate the broad sensitivity of the brain to ambiguous inputs in many modalities, and would suggest that the brain may represent and resolve between the multiple possibilities generated by ambiguous inputs in a similar way, even when such ambiguities occur at very different levels of abstraction.

The brain must constantly interpret an overwhelming but also limited and ambiguous amount of sensory input in order to generate our subjectively rich phenomenological experience of the world, and the ability to create coherence out of the ambiguity is essential in this process. Such resolution of ambiguity factors into models of perception such as Predictive Coding Theory, in which the brain is conceptualized as constantly generating predictions and evaluating the reliability of stimulus inputs in order to construct our perceptions, theories of language comprehension such as the constraint-based model with its simultaneous parallel processing of multiple interpretations of a sentence, and into theories of consciousness, such as Higher-Order Thought theory, that propose that consciousness itself may function as or be dependent upon mechanisms of reality monitoring. The present finding that complex information such as conceptual representations of sentences may also behave bistably in the brain opens the possibility that other levels of representation and abstraction may behave in similar ways, and might ultimately inform our understanding of the time course of how and when the uncertain representations constantly being constructed in our minds make the jump from uncertainty into subjective reality.

Bibliography

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

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, *14*(2), 201–212.

Britz, J., & Pitts, M. A. (2011). Perceptual reversals during binocular rivalry: ERP components and their concomitant source differences: Perceptual reversals during binocular rivalry. *Psychophysiology*, *48*(11), 1490–1499.

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, *47*(4), 474–478.

Davidson, G. D., & Pitts, M. A. (2014). Auditory event-related potentials associated with perceptual reversals of bistable pitch motion. *Frontiers in Human Neuroscience*, *8*.

Debruille, J. B., Brodeur, M. B., & Franco Porras, C. (2012). N300 and Social Affordances: A Study with a Real Person and a Dummy as Stimuli. *PLoS ONE*, *7*(10), e47922.

Farah, M. J., Weisberg, L. L., Monheit, M., & Peronnet, F. (1989). Brain Activity Underlying Mental Imagery: Event-related Potentials During Mental Image Generation. *Journal of Cognitive Neuroscience*, *1*(4), 302–316.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.

Fulford, J., Milton, F., Salas, D., Smith, A., Simler, A., Winlove, C., & Zeman, A. (2017). The neural correlates of visual imagery vividness – An fMRI study and literature review. *Cortex*.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226–241.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *113*(8), 1339–1350.

Intaitė, M., Koivisto, M., & Revonsuo, A. (2013). Perceptual reversals of Necker stimuli during intermittent presentation with limited attentional resources: Perceptual reversals under perceptual load. *Psychophysiology*, *50*(1), 82–96.

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.

Jimenez-Wieneke, A., and M. Pitts. “Neural correlates of expected and unexpected perceptual transitions of a bistable figure.” Reed College Thesis (May 2017)

Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification: Comprehending visual context influences. *Psychophysiology*, *48*(4), 495–506.

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., & Bach, M. (2014). EEG Correlates of Perceptual Reversals in Boring's Ambiguous Old/Young Woman Stimulus. *Perception*, *43*(9), 950–962.

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., Hein, C. M., & Bach, M. (2009). Multistable perception: When bottom-up and top-down coincide. *Brain and Cognition*, *69*(1), 138–147.

Kornmeier, J., Pfaffle, M., & Bach, M. (2011). Necker cube: Stimulus-related (low-level) and percept-related (high-level) EEG signatures early in occipital cortex. *Journal of Vision*, *11*(9), 12–12.

Kornmeier, J., Wörner, R., & Bach, M. (2016). Can I trust in what I see? EEG evidence for a cognitive evaluation of perceptual constructs: Can I trust in what I see? *Psychophysiology*, *53*(10), 1507–1523.

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*(1), 621–647.

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

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological* Review, *101*(4), 676–703.

Maguire, M. J., Magnon, G., Ogiela, D. A., Egbert, R., & Sides, L. (2013). The N300 ERP component reveals developmental changes in object and action identification. *Developmental Cognitive Neuroscience*, *5*, 1–9.

Marks, D. F. (1973). VISUAL IMAGERY DIFFERENCES IN THE RECALL OF PICTURES. *British Journal of Psychology*, *64*(1), 17–24.

Mazerolle, E. L., D'Arcy, R. C. N., Marchand, Y., & Bolster, R. B. (2007). ERP assessment of functional status in the temporal lobe: Examining spatiotemporal correlates of object recognition. *International Journal of Psychophysiology*, *66*(1), 81–92.

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, *4*(1), 15–22.

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.

Pitts, M. A., Martínez, A., Stalmaster, C., Nerger, J. L., & Hillyard, S. A. (2009). Neural generators of ERPs linked with Necker cube reversals. *Psychophysiology*, *46*(4), 694–702.

Pitts, M. A., Nerger, J. L., & Davis, T. J. R. (2007). Electrophysiological correlates of perceptual reversals for three different types of multistable images. *Journal of Vision*, *7*(1), 6.

Renoult, L., Wang, X., Calcagno, V., Prévost, M., & Debruille, J. B. (2012). From N400 to N300: Variations in the timing of semantic processing with repetition. *NeuroImage*, *61*(1), 206–215.

Sedivy, J. (2014). *Language in mind: an* introduction *to psycholinguistics* (First Edition). Sunderland, MA: Sinauer Associates, Inc.

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.

Vissers, C. T. W. M., Kolk, H. H. J., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: Evidence from ERPs in a picture–sentence matching task. *Neuropsychologia*, *46*(4), 967–982.

Zhou, W., & Chen, D. (2009). Binaral Rivalry between the Nostrils and in the Cortex. *Current Biology*, *19*(18), 1561–1565.