INVESTIGATING LOCAL AND CONFIGURAL SHAPE PROCESSING WITH STEADY STATE VISUAL EVOKED POTENTIALS

SHAYA SAMET

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

AUGUST, 2025

Abstract

Object recognition relies on shape, comprised of both local and configural shape information. Local shape information involves elements such as line curvature and edge angles confined to specific regions and perceived independently, while configural shape information arises from the spatial arrangement of local shape features. Previous research suggests that humans primarily use edges and contours for object recognition and are sensitive to disruptions in configural shape. However, it remains unclear how the human visual system separately encodes local and configural shape information, and whether distinct neural mechanisms underlie these processes. We presented stimuli manipulating local and configural shape information independently to participants while recording EEG responses using a Steady-State Visual Evoked Potentials paradigm. Our results reveal that object recognition depends on two mechanisms: one more transient that is sensitive to local shape localized in occipital regions, and one sustained and sensitive to configural shape manipulations localized in right temporal cortex.

Acknowledgments

To my wife, Mira. Mira, words cannot express my gratitude for your unwavering support throughout my academic journey. You have stood by me through every late night, every stressful deadline, and every triumph and setback. As the mother of our four wonderful children, you balance the demands of family life with incredible strength, and I am truly blessed to have you as my wife and best friend. Thank you for your dedication as a nurse and for working tirelessly to support our family. Most of all, thank you for your love, patience, and belief in me. This accomplishment is as much yours as it is mine.

To my children. Each of you brings immeasurable joy and purpose to my life: Evie, you are curious, caring, eager to learn, head-strong, and act as a second mom by always looking for ways to help care for your siblings. Tzvi, my curious and fun-loving boy, you can bring a smile to anyone's face and are fiercely protective, always ensuring your siblings stay safe. Shiri, my sweet and adorably stubborn little girl, you have always known what you want and are not afraid to tell us. When you do you are effortlessly polite, yet still so mischievous. And Gavi, my adventurous, mischievous, friendly little boy, you are a source of constant energy and delight, and at times downright crazy — keeping us all on our toes in the best way possible. You all fill my life with meaning and happiness, and I dedicate this work to you.

To my supervisor, Peter Jes Kohler. Peter, I am profoundly grateful for your mentorship and guidance during my studies, both as an undergraduate and graduate student. You have taught me not only the intricacies of research but also how to grow as a scientist and as a person. Your generosity in providing opportunities, whether through local or international conferences, collaborative projects, or thoughtful feedback, has shaped my academic career in ways I could not have imagined. Thank you for your understanding and flexibility, especially when my family

responsibilities or religious commitments required my attention. Your support has made it possible for me to pursue this work while staying true to my values and priorities.

To my parents, I am grateful for the support you have given me, whether financially through my undergraduate degree, or practically with our children as perfect grandparents. Thank you for your belief in me and for instilling in me the values of hard work, curiosity, and perseverance. Your guidance and love have shaped the person I am today and made this journey possible.

To my siblings, nieces and nephews. Malkie, Berish, Moishy, Hindy, Shloimy, Tzivia, Chavi, Moshe, Yitzi, Shlomo; Shimshy, Sheeree, Aryeh, Yaacov, Tzvi, Daniel; Eli, Chana, Tzvi, Chavi, Huvie, Shmuli, Abie; Johnny, Shoshana, Zoe, Jakey; Chaim, Shevi, Tzvi, Yael, Benny, Levi; Shmuli, Rachel, Abie, Benny, Nava; Dovid, Sara, Rivka; Nosson, Ayala, Tehila, Aliza; and Ben—I'm lucky to have you all in my life. Thank you for the love, laughter, and occasional chaos you bring. I'm proud to be your brother and "favorite" uncle.

To Judah, your excitement about my work and all your practical tips for getting into research and navigating York have been invaluable. You are a brilliant scientist I will always look up to. Thank you, brother.

To my collaborators. Erez Freud, James H. Elder and Nick Baker—thank you for your insights, guidance, and the stimulating discussions that have shaped this work. I've learned so much from each of you, and I'm grateful for your contributions, support, and the spirit of collaboration that made this research both comprehensive and enjoyable.

Thank you to my KVNL lab mates, both present and former—Yara, Shenoa, Denis, Mohammad, Sara, Shadi, Isi, Alejandro, Nikan, Sam, Aurore, Chi, and Jasman—as well as to my cohort and friends at York University—Remy, Tasfia, Zoha, Ricky, Maria, Mario, Nathanial,

Felicia, Emily, and Zack. A special shoutout goes to Jasman for her incredible effort in collecting so much of my data—I truly couldn't have done this without you. Thank you all for the camaraderie, conversations, technical help, and the many laughs along the way. Working alongside you has made this journey far more enjoyable and rewarding.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Human Visual System

Object recognition is supported by a series of spatially organized brain regions that transform retinal input into increasingly abstract and meaningful representations. Visual processing begins in the retina, where light is transduced by photoreceptors and processed by distinct classes of retinal ganglion cells (RGCs). Among these, parasol cells are sensitive to motion and coarse luminance changes, while midget cells respond to fine spatial detail and color. The axons of these RGCs converge at the optic disc to form the optic nerve, transmitting information toward the brain (Blumenfeld, 2022; Prasad & Galetta, 2011).

At the optic chiasm, axons from the nasal hemiretina of each eye cross to the opposite hemisphere, forming the optic tracts. Each tract thus carries visual information from the contralateral visual field, integrating input from both eyes. Most optic tract fibers terminate in the lateral geniculate nucleus (LGN) of the thalamus, a key relay station composed of six stratified layers organized into magnocellular, parvocellular, and koniocellular subdivisions. These layers process distinct aspects of the visual scene: motion and luminance (magnocellular), fine detail and color (parvocellular), and additional chromatic and integrative information (koniocellular) (Blumenfeld, 2022; Prasad & Galetta, 2011).

From the LGN, visual information is projected via the optic radiations to the primary visual cortex (V1) in the occipital lobe. V1 is also known as the striate cortex, derived from the Latin word striatus, meaning “striped” because of its distinctive striped appearance under the microscope, caused by the dense band of myelinated fibers (the line of Gennari) running through it (Hubel & Wiesel, 1977). This projection follows two major anatomical routes: The retrolenticular segment, which ascends through the parietal lobe to the superior bank of the striate

cortex, and Meyer's loop, which curves through the temporal lobe and terminates in the inferior bank. Beyond V1 lies the extrastriate cortex, which encompasses visual regions outside the striate cortex, including areas such as V2, V3, V4, and other regions along the occipital and temporal lobes. V1 is responsible for encoding basic features such as edges, contours, and orientation—a foundational discovery made by Hubel and Wiesel (1962) in their seminal work on simple and complex cells in the early visual cortex. V1 and V2 are now known to be sensitive to contrast, orientation, and spatial frequency (De Valois & De Valois, 1990; Mazer et al., 2002).

As visual processing advances through the extrastriate cortex, information is distributed into two major cortical processing streams: the dorsal stream and the ventral stream. The dorsal stream supports spatial localization and action planning, while the ventral stream supports object identity recognition. Based on this functional distinction, the dorsal and ventral pathway are commonly referred to as the "where" and "what" pathways, respectively (Ungerleider & Mishkin, 1982; Goodale & Milner, 1992). Within the ventral stream, area V4 plays a pivotal role in bridging low-level feature encoding and high-level object representation. Neurons in V4 exhibit selectivity for complex shape features, including curvature, angular fragments, and specific contour combinations (Pasupathy & Connor, 1999; Hegde & Van Essen, 2001; Brincat & Connor, 2007). Rather than simply responding to oriented edges, many V4 neurons integrate local contour fragments into intermediate-level shape representations, encoding both the form and the spatial arrangement of visual elements (Gallant et al., 1993; 1996). Additionally, V4 contributes to color processing, maintaining shape selectivity even when stimuli differ only in color but not in brightness, where luminance cues are absent (Bushnell et al., 2013). This dual sensitivity to both form and color suggests that V4 functions as a convergence zone for diverse visual features,

providing a crucial substrate for constructing robust object representations (Roe et al., 2012; Willmore et al., 2010).

Beyond V4, ventral stream regions such as the lateral occipital complex (LOC) and inferotemporal cortex (IT) progressively integrate these intermediate shape signals into more complex representations of objects (Rust & DiCarlo, 2010). In these higher-level areas, neurons exhibit increasing tolerance to transformations like position, size, and viewpoint, while maintaining high selectivity for specific object identities. Together, this hierarchical processing architecture enables the visual system to transform local features into coherent object percepts, allowing recognition under diverse and challenging conditions. For example, LOC responds robustly to objects, while more anterior regions encode configural properties and semantic identity of visual stimuli (Kourtzi & Kanwisher, 2001; Cant & Goodale, 2007).

As visual signals ascend the ventral stream, they are transformed into representations that support object recognition (Kourtzi & Kanwisher, 2001; Cant & Goodale, 2007; Grill-Spector et al., 2001). Electrophysiological and neuroimaging evidence suggests that the brain first extracts coarse visual information rapidly, likely reflecting early stages of processing, as early components of steady-state visual evoked potentials (SSVEPs) reflect low-level visual properties, while sustained activity over longer time intervals is associated with higher-level perceptual integration and object-specific representation (Ales et al., 2012; Liu-Shuang et al., 2014). Furthermore, neural responses linked to detecting symmetry, a mid-level structural property, emerge at later latencies compared to responses for simple visual features, suggesting that more complex representations emerge later, requiring sustained and possibly recurrent computations (Kohler et al., 2016). These later stages enable the visual system to recognize and differentiate objects under conditions of variation and ambiguity, a capacity that depends critically on how visual shape is represented.

This anatomical and functional architecture enables the brain to integrate local features (e.g., curvature, edges) and configurations (e.g., spatial relationships between parts), allowing for robust and rapid object recognition under a wide range of viewing conditions. The current study investigates how two types of shape information, local and configural, are differentially processed across cortical regions and over time.

1.2 Shape Perception: Local, Configural and Global Processing

When objects are viewed from different angles, the retinal image of an object also changes, but our perception remains stable due to the brain's ability to interpret these changes in light of the object's three-dimensional structure (Palmer, 1999). Research shows that edge-based representations, rather than surface characteristics such as texture, are critical for real-time object recognition (Biederman & Ju, 1988). Elder and Velisavljević (2009) further demonstrated that shape is a primary cue in the rapid detection of objects, particularly animals, in natural scenes.

While shape constancy allows for perceptual stability across changing viewpoints, recognition performance is not entirely viewpoint-invariant. Palmer, Rosch, and Chase (1981) showed that objects are recognized more quickly and accurately when viewed from canonical perspectives—typically orientations that provide the most structurally informative view of the object. For example, a three-quarter view may reveal both the front and side surfaces, conveying more spatial structure than a profile or head-on view. These canonical perspectives align with how objects are typically encountered and interpreted in the real world. As a result, changes in orientation can systematically alter how shape is processed sometimes making even familiar objects difficult to recognize. This introduces an important caveat to shape constancy: while the

visual system can compensate for viewpoint variation, there are optimal viewpoints that facilitate access to object structure more effectively than others.

The shape of an object is comprised of local, global and configural shape. Local shape information refers to curvature or the angle between edges at specific regions along a shape contour (Pasupathy & Connor, 1999; Elder & Zucker, 1998). Local shape can often be perceived independent of the object as a whole and can be combined by the visual system to form the basis of global and configural shape representations, supporting perception of the object as a whole. Global shape refers to the overall contour or silhouette of the object – its outer boundary that defines the object as a whole, without regard to internal relations among features (Elder & Velisavljević, 2009; Baker et al., 2018). Configural shape, in contrast, refers to the spatial organization and relationships among multiple parts or features within the object, emphasizing how parts are positioned relative to each other to form a coherent whole (Pomerantz, 1986; Biederman, 1987; Palmer, 1999). Thus, global shape captures the big-picture of the object's outline, while configural shape reflects the structured arrangement of its internal components.

The importance of configural shape processing in object recognition is well established. Prior work shows that spatial relations among parts can enhance perceptual salience, making wholes more recognizable than isolated components—a phenomenon known as configural superiority effects (Pomerantz, 1986). Objects are recognized not only by local features but also by the spatial relationships among parts, as in recognition-by-components theory, which proposes that objects are identified through configurations of simple volumetric parts (Biederman, 1987). Relational information has also been shown to contribute to perceiving coherent forms (Palmer, 1999). Consistent with these findings, Baker and Elder (2022) demonstrated that disrupting configural relationships impairs object recognition using silhouette stimuli, which remove internal

features and isolate shape-based cues. They manipulated the silhouettes by dividing the object horizontally, rotating one half and reattaching it, thereby disrupting the configural relationships among parts. This manipulation altered the overall object shape while preserving many of the low-level visual features such as edge continuity and complexity. Human recognition performance declined significantly for these reconfigured silhouettes, suggesting that configural shape is important for object recognition.

1.3 Effects of Inversion on Shape Perception

Evidence in favor of the importance of configural shape in object recognition has also come from inversion effects, which disrupts our ability to perceive configural shape (Palmer, 1999). Inversion effects have been consistently observed across a variety of object categories. For example, Jolicoeur (1985) and Lawson and Humphreys (1996) found that recognition of everyday objects is hindered by inversion, especially when objects have a canonical orientation or when recognition depends on viewpoint-specific matching. Similarly, car experts experience strong inversion effects for cars, suggesting that orientation sensitivity can be modulated by experience and category-specific expertise. These findings underscore that orientation is a critical factor in object recognition (Rossion & Curran, 2010).

Faces have consistently been found to show strong inversion effects. Yin (1969) showed that inversion disproportionately affects faces compared to houses, airplanes, and people, likely due to faces being perceived as a whole and therefore relying more on configural processing compared to the feature-based processing used for other objects, establishing what is now known as the face inversion effect. This finding provided early evidence that faces are processed differently than other objects, specifically through mechanisms attuned to upright, configural

information. Importantly, however, the face inversion effect should be distinguished from the composite face effect (Young et al., 1987). While inversion effects demonstrate a general orientation sensitivity, the composite face effect provides more specific evidence for configural processing, as observers are impaired at recognizing identical facial halves when aligned with incongruent counterparts, a phenomenon not seen in non-face objects. Together, inversion and composite effects converge to highlight the importance of configural processing for object recognition.

Observers are also better at recognizing individual facial features (nose or eyes) when those features are presented within the context of a whole face, rather than shown in isolation (Tanaka & Farah, 1993). This part-whole effect was not observed for non-face objects like houses, where isolated features were recognized just as well as when embedded in the whole. This finding suggests that faces are processed as integrated wholes, and that their recognition depends on more than the sum of their parts. These findings provide support for the notion that face recognition depends uniquely on configural mechanisms, in contrast to the more part-based processing used for other object categories.

However, recent research has challenged the idea that inversion effects are a direct and exclusive signature of configural processing. Gerlach (2025) found large inversion effects across a range of common objects, including those from structurally similar categories that are thought to require little or no configural integration. These objects, such as cups or chairs, share consistent overall structure, making them typically identifiable through local features. This suggests that inversion effects can arise even in the absence of strong configural demands, and may reflect more general disruptions to visual processing, such as reduced familiarity with inverted forms or impaired access to stored object representations. Yet, configural structure clearly still plays an

important role in modulating inversion sensitivity. Brandman and Yovel (2012) showed that inversion effects persisted for faceless heads when minimal body context preserved global shape but disappeared when configural cues were further degraded (e.g., headless bodies or isolated heads without context). Similarly, Baker and Elder (2022) demonstrated that disrupting configural relationships in silhouettes through a "Frankenstein" manipulation reduced the magnitude of inversion effects, indicating that while inversion is not solely tied to configural processing, its impact is modulated by the presence of configural shape cues.

In sum, inversion effects appear to be driven largely by disruptions to configural shape processing. Although some evidence suggests that local shape cues can also contribute, studies consistently show that preserving or disrupting configural relationships strongly modulates inversion sensitivity, highlighting the central role of configural shape in object recognition. It is important to distinguish between inversion as a general index of orientation and familiarity effects, and inversion as evidence of configural processing. While the composite face effect (Young et al., 1987) provides stronger evidence for specifically configural mechanisms in face recognition, the inversion effects discussed here may be interpreted as reflecting more general orientation sensitivity in configural shape processing.

1.4 Artificial Neural Networks and Human Object Recognition

In recent years, artificial neural networks (ANNs), particularly convolutional neural networks (CNNs) or deep convolutional neural networks (DCNNs), have become prominent in computer vision for image recognition tasks. These models have been shown to predict neural responses in object-selective regions of the primate ventral visual stream (Cadieu et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2014). However, there are notable differences between DCNNs and humans in the visual features they use for identifying shapes. As

discussed in the previous section, humans rely heavily on global shape as a cue for object recognition (Biederman & Ju, 1988; Elder & Velisavljević, 2009), while DCNNs tend to focus on local features such as edges as well as textures. Baker et al. (2018) demonstrated that DCNNs are sensitive to local contour information but do not represent the spatial relationships between parts.

As discussed above, Baker and Elder (2022) demonstrated that configural shape information is important for object recognition in humans. The same study also demonstrated that DCNNs lack the configural-relationship sensitivity characteristic of human shape perception. Interestingly, both humans and DCNNs exhibit performance drops in making perceptual judgments when objects are inverted. In fact, DCNNs are affected even more than humans (Baker & Elder, 2022). This may be because the DCNNs solve the task using local shape information, i.e. curvature at specific image locations, which is heavily disrupted by inversion. Humans, in contrast, rely on global and configural relationships, which while also disrupted by inversion, may be able to overcome it to some extent. These results show that the human sensitivity to spatial relationships among object parts is not yet replicated in artificial systems. A related distinction was demonstrated by Baker et al. (2018) and Baker et al. (2020) who found that while humans rely on global and/or configural shape for object recognition, DCNNs instead classify objects based on local contours fragments and textures, struggling to recognize silhouettes when local contours are disrupted and often misidentifying stimuli like circles made of corners or squares made of curves when global shape cues are preserved. Similarly, Geirhos et al. (2019) found that standard ImageNet-trained CNNs exhibit a strong bias toward local texture rather than global shape, in contrast to human perception. By increasing shape-based inductive biases through training on stylized images, CNNs became more robust and human-like in their recognition strategies. These

results show that while humans rely mostly on configural shape, ANNs rely on local features like curvature.

Building on the distinction between local and configural shape processing, and prior findings that inversion disrupts configural sensitivity, we investigated how the human brain processes local and configural shape information, providing insights into how these different aspects of shape are engaged during object recognition. We specifically asked whether the human visual system shows distinct neural sensitivity to manipulations of local and configural shape information, and how this sensitivity is influenced by stimulus inversion. This line of investigation provides a foundation for understanding how the brain processes different aspects of shape, and how this may contrast with the processing strategies employed by ANNs. We created three studies designed to manipulate local and configural shape, with the following overarching hypotheses:

(1) Disruptions to local curvature structure would lead to measurable changes in neural responses, with increasing local similarity to natural shapes producing correspondingly stronger responses.
(2) Disruptions to configural shape—such as altered spatial relationships among parts—would lead to differing neural responses, reflecting the brain's sensitivity to configural shape organization.
(3) Inversion would attenuate, but not eliminate, sensitivity to configural shape, consistent with prior findings that orientation impacts configural shape perception more strongly than local detail.

To test these hypotheses, we employed a Steady-State Visual Evoked Potentials (SSVEPs) EEG paradigm. This method allows us to isolate neural responses to specific visual features by analyzing frequency-tagged signals in EEG data, making it ideal for examining the distinct brain responses to shape manipulations. Participants passively viewed stimuli that selectively manipulated local (Curvature Matched Controls) or configural (Frankenstein Shapes) shape information, alongside animal silhouettes. By presenting upright and inverted versions of each stimulus, we examined the extent to which orientation modulates configural shape processing in the human brain.

2. Methods

2.1 Participants

Participants were undergraduate students recruited from York University's Undergraduate Research Participation Pool (URPP). In exchange for participation, students received 2.5 research credits applicable to their grade in PSYC 1010.6.0: Introduction to Psychology. All participants provided informed consent and had normal or corrected-to-normal vision as confirmed using a Snellen chart.

- **Experiment 1** included 35 participants (25 female, 10 male) ranging in age from 18 to 66 years (M = 22.83, SD = 7.78). Data from 32 participants were included in the final analysis.
- **Experiment 2** included 29 participants (21 female, 8 male) aged 17 to 37 years (M = 19.68, SD = 3.72). Data from 28 participants were included in the final analysis.
- **Experiment 3** included 31 participants (26 female, 5 male) aged 19 to 43 years (M = 21.87, SD = 5.02). Data from 30 participants were included in the analysis.

All procedures were approved by York University's Research Ethics Board. Informed consent was obtained from each participant prior to the study. Personal information was collected and securely stored using REDCap: Research Electronic Data Capture tools (Harris et al., 2009), and all data was anonymized to maintain confidentiality. Participants were informed of their right to withdraw at any time without penalty. Individuals with a history of epilepsy or sensitivity to flashing lights were excluded from participation to mitigate any risk associated with visual stimulation. No adverse events were reported during data collection.

2.2 Stimuli

For our studies, we used two general classes of stimuli. The first class, Curvature Matched Controls (CMC), was taken from a stimulus set developed by Elder and colleagues (2016, 2018; see Figure 1). These stimuli are statistically controlled, abstract shapes that do not contain configural shape information consistent with any natural shapes, but are matched in terms of local contour information. The shapes were generated by using a maximum entropy algorithm trained on curvature statistics derived from the Hemera dataset of animal silhouettes, originally collected by photographing animal models against a green screen and extracting their outlines. Each CMC is represented as a closed equilateral polygon (typically with 120 sides), where the shape is fully specified by a sequence of turning angles between adjacent edges. A Markov chain Monte Carlo (MCMC) method was used to sample from shape spaces that match progressively more detailed statistical properties of natural curvature distributions. These include the mean, variance, skewness, kurtosis, and higher-order correlations of local curvature (Figure 2). At the lowest constraint level only the mean of curvature distribution is matched; additional constraint levels successively match variance (spread of curvature values), skewness (asymmetry in the curvature distribution), kurtosis (reflecting the prevalence of extreme curvature values), as well as higher-

order moments. Thus, although each CMC forms a coherent global outline, it lacks the naturalistic configural organization consistent with any recognizable shape silhouettes. Because these stimuli control local curvature statistics, they are ideal for assessing the visual system's sensitivity to local shape information in isolation. For our studies, we used four of the five curvature-matched control stimuli categories: *no constraints* (only mean is matched), *variance* (mean and variance is matched), *varkurt* (mean, variance, skewness and kurtosis is matched), all (the curvature distribution is fully matched, including higher-order moments).

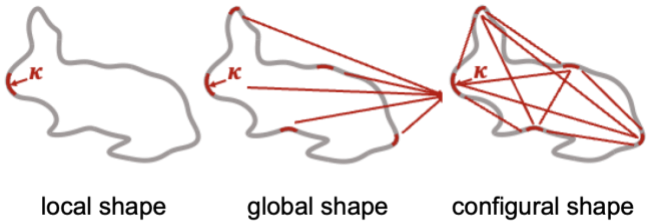

Figure 1: Local, Global and Configural Shape

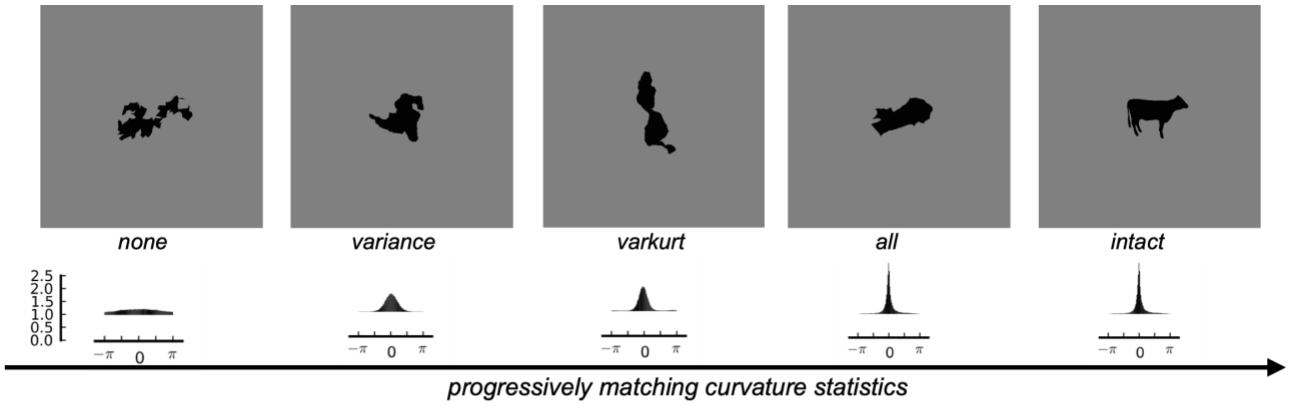

Figure 2: Maximum Entropy Shape Curvature-Matched Controls.
Curvature statistics from natural shapes were extracted, including mean, variance (spread of curvature values), kurtosis (frequency of extreme curvature values), as well as higher-order moments. Synthetic shapes were then generated to progressively match these statistical properties, resulting in stimuli that align with natural shape curvature at increasing levels of complexity. This method allows for precise manipulation of local curvature statistics, enabling analysis of how each statistical property influences neural responses to shape. These curvature-matched controls were used for studies 1 and 2.

The second class of stimuli used for our studies were so-called "Frankenstein" shapes (Baker & Elder, 2022) (Figure 3). The stimuli were animal silhouettes whose configural relationships are manipulated while preserving local shape. In their paper, Baker and Elder (2022) introduced two manipulated stimulus sets: Fragmented and Frankenstein. In the fragmented

condition, the top half of the silhouette was flipped horizontally and displaced laterally, resulting in two abutting but disjointed parts. In the Frankenstein condition, the flipped top half was then realigned vertically with the bottom half, preserving the silhouette as a single object while still disrupting the spatial relationship between its upper and lower parts (see Figure 3). For the present study, we focused exclusively on the Frankenstein condition, as it maintains the object's continuity while selectively targeting configural shape. We used these stimuli to measure the sensitivity of the visual system to configural shape information. To further tie our measurements to behaviour, we also made a distinction between animal shape silhouettes for which the Frankenstein manipulation has a strong effect on object recognition, and those for which the effect is weak or absent. This distinction was based on a 9-alternative forced choice (9AFC) recognition task comparing performance on Frankenstein versus intact versions of the same exemplar. Exemplars were classified as "strong" when the manipulation produced a 23–50% drop in recognition accuracy (M = 33.9%, SD = 8.3%), and as "weak" when the drop was minimal (–0.67% to 10%, M = 1.4%, SD = 5.4%) (see Figures 3–4). Finally, we also used both upright and inverted versions of animal silhouettes to measure inversion effects.

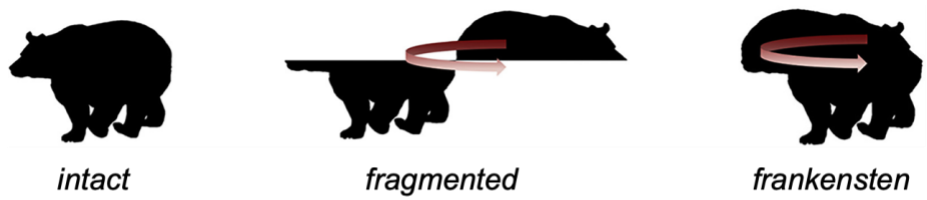

Figure 3: Frankenstein Shapes.
An animal shape is cut in half horizontally, rotated 180 degrees and reattached, disrupting configural shape relationships.

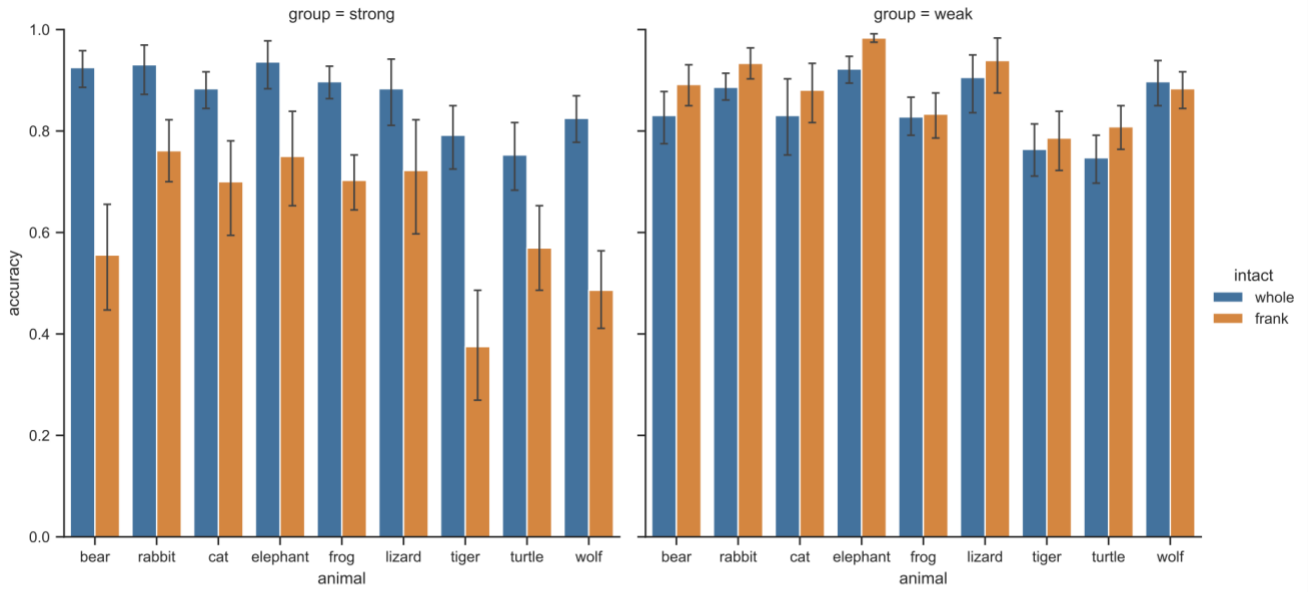

Figure 4: Behavioural Recognition Data.
Participants were presented both Frankensteined and intact animal shapes for 150ms followed by a mask, participants were then asked to categorize the animal in a 9 alternative forced-choice task. Frankenstein shapes were classified as strong and weak based on how much recognition performance dropped compared to intact versions.

2.3 Procedure

We employed a Steady-State Visual Evoked Potentials (SSVEPs) paradigm using EEG to measure neural responses to shape stimuli. SSVEPs rely on the brain's consistent oscillatory responses to periodically presented stimuli, generating frequency-tagged signals that can be analyzed in the frequency domain (Norcia et al., 2015). This approach allows us to isolate neural

responses to specific shape manipulations while minimizing the influence of confounding factors like attention, as participants were instructed to passively fixate on a central fixation cross throughout 180 trials.

Each trial consisted of 12 one-second cycles, with each cycle presenting pairs of images sequentially, each image shown for 500 ms. This gave rise to a stimulation frequency of 1 Hz. The first and last cycles were excluded from analysis, leaving 10 cycles for further analysis. This was done to ensure that only data for which participants were fixating and fully engaged in the stimuli were included in the analysis. In SSVEP paradigms, periodic stimulation produces distinct peaks in the EEG frequency spectrum at integer multiples of the stimulation frequency, known as harmonics. Even harmonics in the EEG signal capture components of the brain response that is the same for the two image types being presented. This tends to reflect general visual processing driven by changes in luminance or low-level visual features that occur for both image types. Odd harmonics, on the other hand, capture components of the brain response that are distinct for the two image types, reflecting brain responses to the shape information that is present in one image type but not the other (Kohler et al., 2016). This allowed us to isolate and compare brain responses associated specifically with differences in local versus configural shape information. This procedure was used for all studies presented here. In each of our three studies, images from different stimulus were paired to measure local and configural shape processing. We will now discuss the specifics of each study.

2.4 Experiment Design

Study 1: Local Shape Sensitivity

For Study 1 we used three CMC shape comparisons as well as a comparison between CMC shapes and intact animal shapes. See above for descriptions of the CMC shapes as well as Figure 5:

- *no constraints (none) vs variance* (Figure 5a)
- *variance vs varkurt* (Figure 5b).
- *varkurt vs all* (Figure 5c)
- *all vs intact upright* (Figure 5d)

This "step-wise" set of conditions allowed us to isolate and incrementally test the contribution of increasing local curvature statistics, and ultimately, the influence of global shape on visual processing. If a given condition elicits measurable odd harmonic responses, it indicates sensitivity to the additional curvature constraint. We also added an inverted version of the *all vs intact* animal comparison (Figure 5e) allowing us to evaluate how inversion affects responses to global shape cues when local statistics are held constant. Any differences in the responses for this comparison can be attributed to configural shape, given that the local shape statistics are identical for both objects.

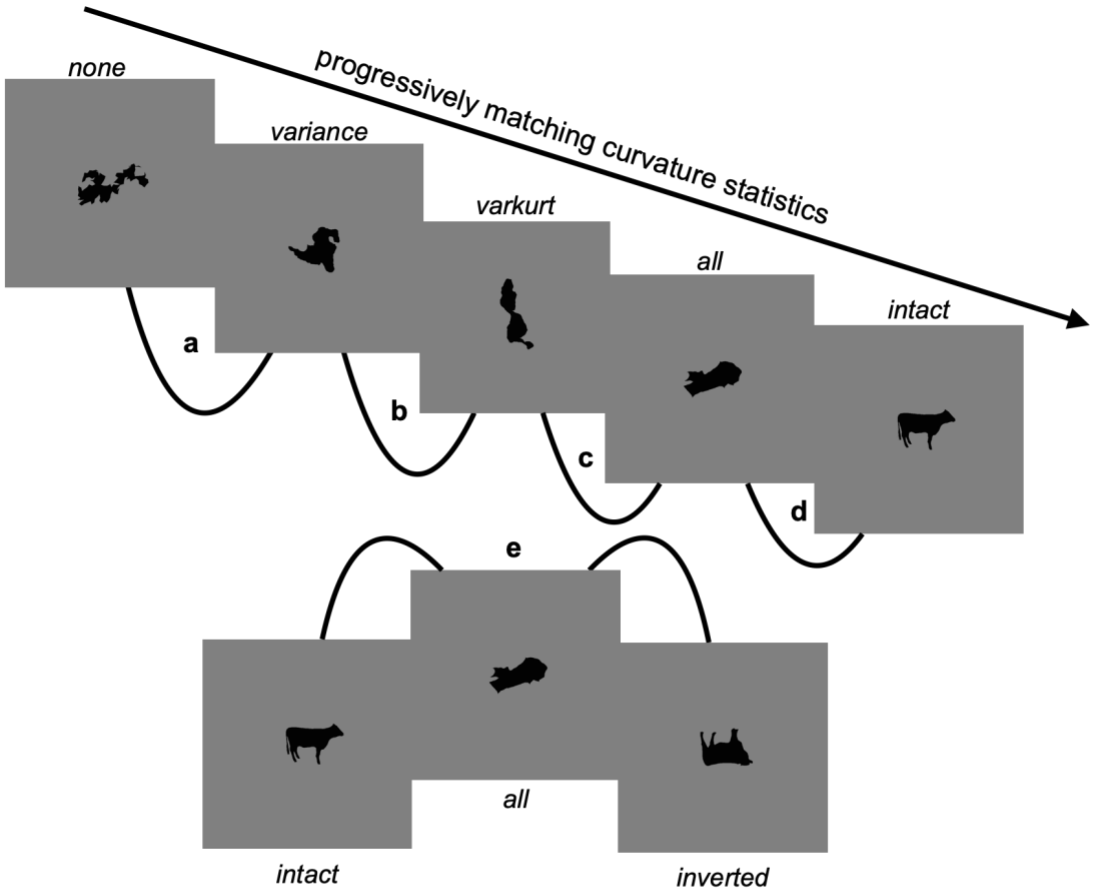

Figure 5: Study 1 comparisons.
The distribution of the curvature statistics is progressively matched until fully aligned with the natural animal shape. Each level of constraints is compared to the subsequent level with increased constraints (a-d), demonstrating the visual system's response to progressively matching the curvature distribution. Additionally, an inverted version of the fully matched versus intact animal comparison is included to assess how inversion affects responses to global shape cues when local curvature statistics are held constant (e).

Study 2: Naturalness Gradient

Study 2 tested whether progressively constraining curvature statistics in CMC stimuli would elicit brain responses that more closely resemble those evoked by natural animal shapes. We paired each level of CMC constraints with intact animal silhouettes (Figure 6). Reduced response amplitude with added constraints suggests greater similarity between the synthetic and natural animal shapes.

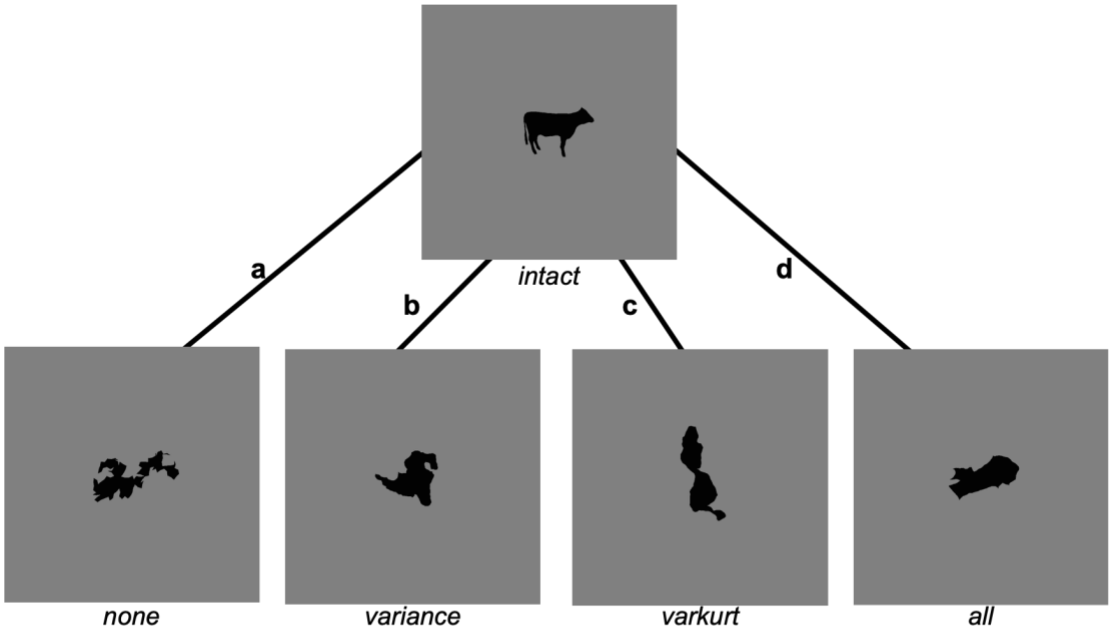

Figure 6: Study 2 comparison.
Comparing each level of constraints with intact shapes to investigate the extent to which matching the curvature distribution produces responses similar to that of natural intact animal shapes.

Study 3: Configural Shape Disruption

Study 3 investigated the impact of disrupting configural shape using Frankenstein stimuli. We compared Frankenstein shapes with intact animals (Figure 7), separately analyzing strong and weak conditions. This allows us to measure the effect of the Frankenstein manipulation on brain responses to configural shape. Our 2 × 2 design (*strong vs weak* × *upright vs inverted*) makes it possible to ask whether these responses are modulated by inversion, and by the size of the behavioral effect. Finally, to explore orientation effects, we also compared inverted Frankenstein shapes with inverted intact animals.

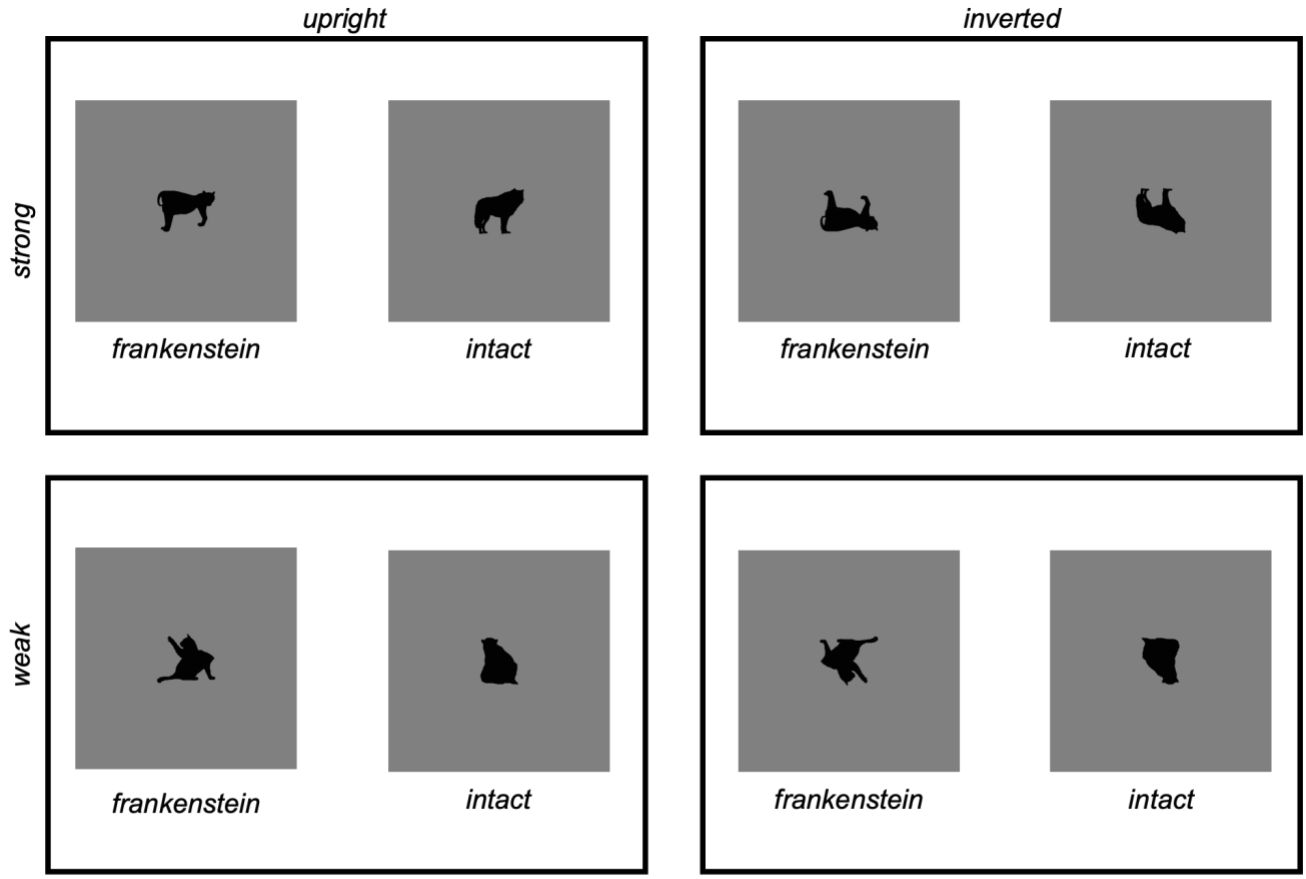

Figure 7: Study 3 comparisons
Study 3 explored the effects of configural disruption on object processing. We ran the same basic comparison, frankenstein vs. intact, for strong and weak frankensteins, and for upright and inverted versions of the stimuli, resulting in a 2 x 2 design.

2.5 Analyses

We defined electrode Regions-of-Interest (ROIs) based on previous studies in the lab, six electrodes were selected over occipital area, and nine electrodes were selected for each of the left and right temporal regions (see Figure 8). As described above, odd harmonics of the stimulation frequency capture neural responses that differentiate between the two alternating images in each pair, reflecting the visual system's sensitivity to shape differences across conditions. Accordingly, we focused our analyses on the first and third harmonics of the stimulation frequency, because

these components consistently yield the strongest and most reliable signals in SSVEP paradigms and offer meaningful interpretability, while higher harmonics rapidly decline in amplitude and often fall near the noise floor (Rossion et al., 2012; Liu-Shuang et al., 2014; Norcia et al., 2015). Ales et al. (2012) demonstrated that the first harmonic of a 3 Hz SSVEP paradigm was most robust over right occipito-temporal regions and was selectively enhanced by face stimuli, suggesting sensitivity to higher-level, sustained configural processing. On the other hand, Norcia et al. (2015) reviewed evidence that higher harmonics, including the third, often reflect nonlinear responses associated with early, transient visual mechanisms in occipital cortex. Furthermore, Kohler et al. (2016) found that the first, third and fifth harmonics carry rotation symmetry information.

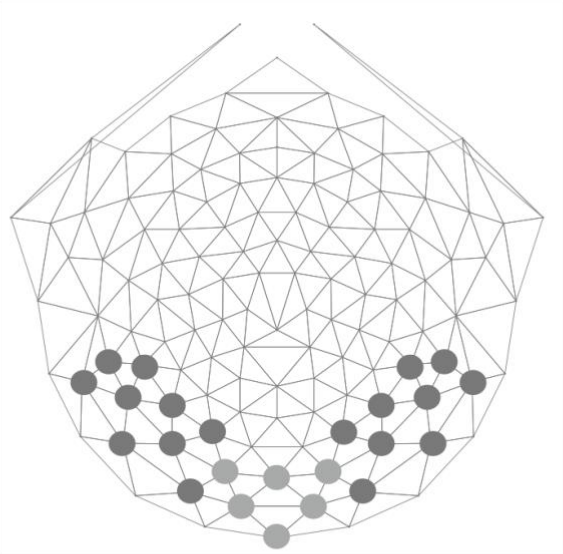

Figure 8: Electrode Regions of Interest.
We selected 9 electrodes each over left temporal and right temporal cortices, and 6 electrodes over occipital cortex.

To quantify steady-state visual evoked potential (SSVEP) responses, we computed group average amplitudes in the frequency domain, using the method as described by Kohler et al. (2018). This was done by first averaging the real and imaginary components of the Fourier-transformed signal across participants, yielding a group vector at each frequency of interest. The

projected amplitude for each participant was then computed as the scalar projection of their individual complex-valued response onto this group vector. This projection captures the component of each participant's response that is phase-aligned with the group average, effectively reducing noise from out-of-phase activity and enhancing signal interpretability. To estimate the noise floor, we averaged the real and imaginary components of the frequency bins immediately adjacent to each harmonic of interest (i.e., the flanking bins), then computed the vector magnitude of that average to estimate a noise floor. This provides a conservative baseline against which the signal at the target frequency can be compared. These projected amplitudes were used for statistical analysis, including repeated measures ANOVAs, and to compute standard errors of the mean for visualization purposes (e.g., error bars in amplitude plots). This approach has been widely used in SSVEP research to retain phase information while enabling group-level inference (e.g., Ales et al., 2012; Cottereau et al., 2011).

To test whether each individual condition produced a measurable response, Hotelling's T^2 tests were used to compare each response to the noise floor, using the real and imaginary components from each participant, and considers both the phase and the amplitude of the response. To assess significance between the different conditions and ROIs, we conducted two-way repeated-measures ANOVAs on projected amplitudes for studies 1 and 2 with factors ROI (three levels) and Condition (four levels). To assess inversion effects in Study 1 we performed a two-way repeated-measures ANOVA on projected amplitudes, with ROI (3 levels) and Orientation (two levels) as within-subject factors. For Study 3 we ran a three-way ANOVA with Orientation (two levels), Strength (two levels) and ROI (three levels) as within-subjects factors. Each study's ANOVA was conducted on the first and third harmonic separately.

3 Results

3.1 Local Shape responses.

First Harmonic

At the first harmonic of Study 1, the results of the two-way ANOVA, with Greenhouse-Geisser corrections applied due to violations of sphericity, revealed a significant main effect of Condition ($F(1.96, 60.76) = 4.98$, $p = 0.010$), indicating differences in response amplitudes across stimulus conditions. The main effect of ROI was not significant ($F(1.94, 60.13) = 2.54$, $p = 0.089$). However, there was a significant ROI × Condition interaction, ($F(3.04, 94.23) = 3.24$, $p = 0.025$). The *all vs. intact* comparison consistently produced the highest response amplitudes across all regions of interest, with the strongest activity localized to the right temporal ROI. The *varkurt vs. all* comparison produced significant responses in the left temporal region, while the *variance* vs. *varkurt* comparison yielded significant responses in the both the occipital region and right temporal region. No other local curvature comparisons reached significance at this harmonic. See Table 1 for all Hotelling's T-tests and Figure 9 for projected amplitudes.

Third Harmonic

For the third harmonic, the ANOVA, with Greenhouse-Geisser corrections applied due to violations of sphericity, revealed a significant main effect of Condition, ($F(2.10, 64.93) = 10.94$, $p < .001$), indicating differences in response amplitudes across conditions. We found no other significant main effects or interactions (ROI: $F(1.56, 48.21) = 1.40$, $p = .254$; ROI × Condition: $F(4.03, 124.80) = 1.92$, $p = .111$.) The *all vs intact upright* comparison produced the strongest overall responses, with the highest amplitudes observed in the occipital ROI and relatively uniform activity across all regions. Additionally, the *none vs variance* comparison yielded significant

responses across all three ROIs, with the largest amplitudes found in the occipital region. Comparisons involving further curvature constraints (e.g., *variance vs. varkurt*, or *varkurt vs all*) did not produce significant differences at the third harmonic (See Figure 9 and Table 1).

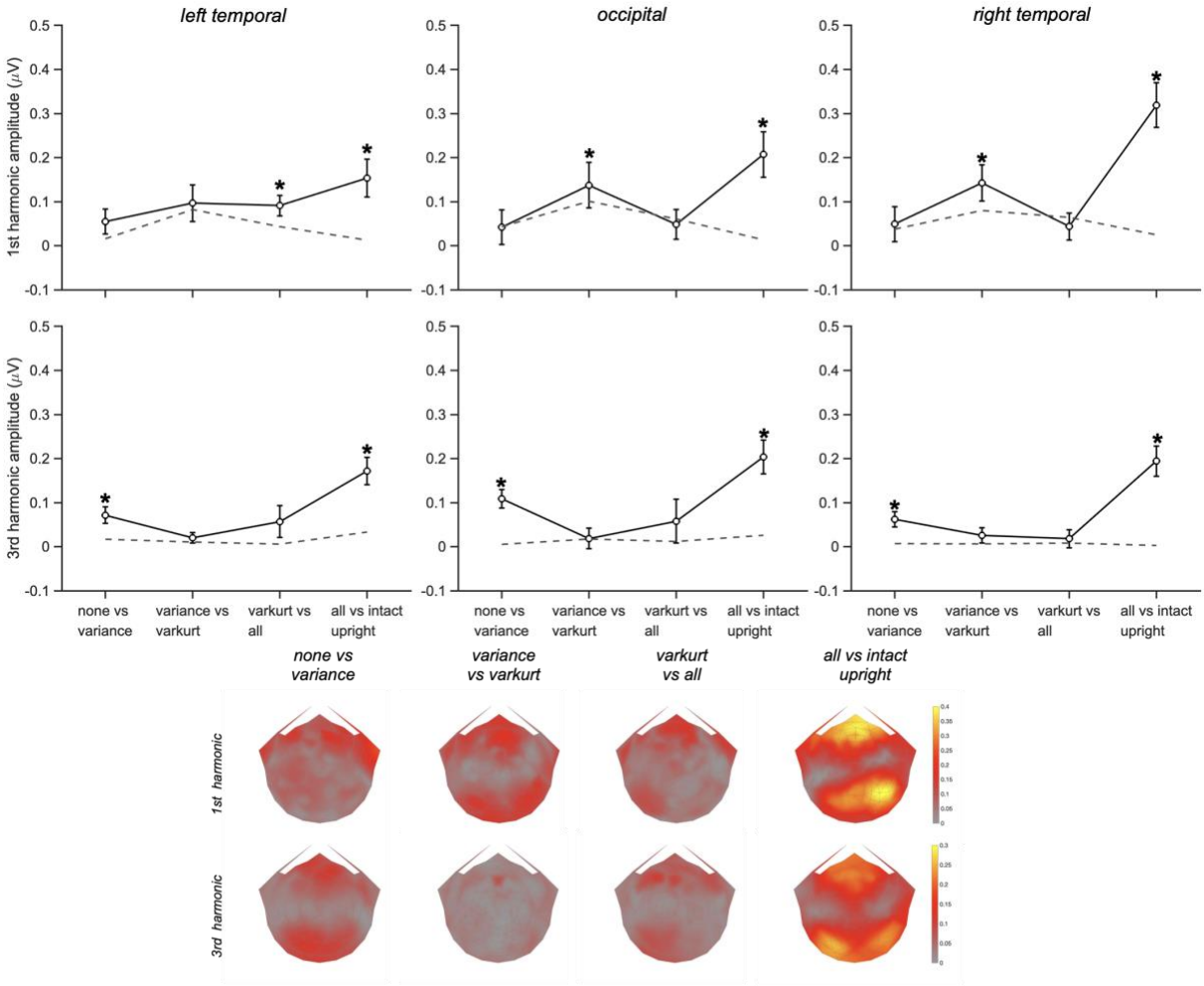

Figure 9: Study 1 Coherent-Averaged Amplitudes Over 3 Electrode Regions of Interest and Topographies.
Group average amplitudes were calculated by separately averaging the real and imaginary components of the EEG data across participants. The magnitude of the group vector was then computed. For each participant, the magnitude of their data's projection onto this group vector was also calculated. Error bars indicate the standard error of the mean (SEM), based on the projected amplitudes. The dashed line is the noise floor, which the projected amplitudes were compared to in order to determine significance via Hotelling's T-tests. Topographic maps display the distribution of these brain responses across the different ROIs, illustrating the spatial distribution of neural activity for each condition. The responses show

	df = 2,31 condition	left temporal		occipital		right temporal	
		p value	T^2-statistic	*p* value	T^2-statistic	*p* value	T^2-statistic
1st harmonic	none vs variance	0.1717	3.8644	0.5376	1.3096	0.4665	1.6165
	variance vs varkurt	0.0545	6.635	0.0318	8.015	0.003	14.6519
	varkurt vs all	0.0023	15.5066	0.3774	2.0808	0.383	2.0484
	all vs intact upright	0.0045	13.4346	< 0.0001	19.7592	< 0.0001	44.1561
	all vs intact inverted	0.4598	1.6525	0.9907	0.0193	0.0876	5.464
3rd harmonic	none vs variance	0.003	14.6851	< 0.0001	27.3653	0.0058	12.7037
	variance vs varkurt	0.2252	3.2397	0.728	0.6631	0.3521	2.2339
	varkurt vs all	0.3055	2.5498	0.4407	1.7405	0.5596	1.2232
	all vs intact upright	< 0.0001	30.9853	< 0.0001	30.8703	< 0.0001	39.7468
	all vs intact inverted	0.001	17.9693	0.0034	14.2762	0.0025	15.2567

Table 1: Study 1 Hotelling's T^2 Tests
Hotelling's T^2 tests comparing each stimulus condition to the noise floor. These tests assess the significance of neural responses to conditions by determining whether the group-averaged project amplitude deviates significantly from the noise. The tests were performed for the first and third harmonics of the EEG signal, and the table includes the corresponding test statistics, degrees of freedom, and p-values. Significant differences indicate robust neural responses to the manipulated shape features.

3.2 Inversion Effects

First Harmonic

The results of the ANOVA revealed a significant main effect of ROI, ($F(2, 62) = 6.49$, $p = .003$), indicating that responses differed across brain regions. There was also a significant main effect of Orientation, ($F(1, 31) = 5.50$, $p = .026$), suggesting overall differences in responses between upright and inverted stimuli. A significant ROI × Orientation interaction emerged, ($F(2, 62) = 9.74$, $p < .001$), reflecting that the effect of inversion differed across regions of interest. At the first harmonic, responses to the *all vs intact upright* comparison were significantly higher

than in the *all vs intact inverted* comparison. In the inverted comparison, amplitudes dropped to the extent that no significant differences were observed. This pattern was most prominent in the right temporal ROI, consistent with the topography seen in the upright condition (see Figure 10 and Table 1).

Third Harmonic

At the third harmonic, the ANOVA revealed a significant main effect of Orientation, ($F(1, 31) = 9.55$, $p = .004$), indicating differences in response amplitude between upright and inverted stimuli. No other significant main effects or interactions were found. While the inversion comparison showed a reduction in amplitude relative to the upright condition, the response remained statistically significant (see Table 1). These responses were strongest in the right temporal region but showed a more uniform distribution across ROIs than at the first harmonic (see Figure 10).

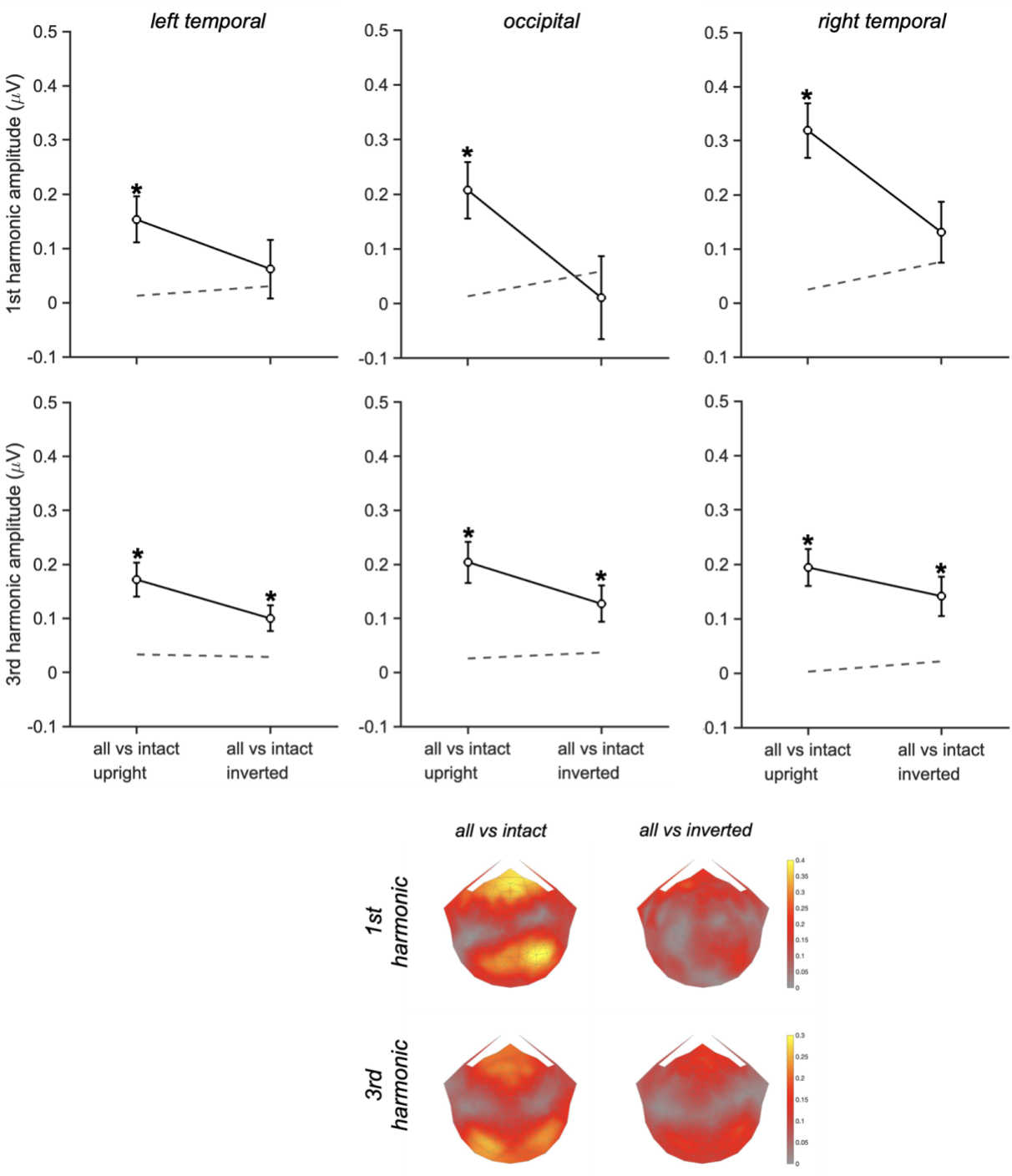

Figure 10: Study 1 Coherent-Averaged Amplitudes Over 3 Electrode Regions of Interest and Topographies.

Figure follows the same principles as Figure 9, dotted line is the noise floor, error bars indicate the standard error of the mean. Topographic maps display distributed brain responses for each condition.

3.3 Study 2

First Harmonic

In Study 2, we examined how similarity in curvature statistics modulates neural responses by comparing each level of curvature-constraints with intact animal shapes. For the first harmonic the ANOVA with Greenhouse-Geisser corrections applied due to violations of sphericity, revealed a significant main effect of ROI, ($F(1.23, 33.23) = 6.15$, $p = .014$), indicating differences in response amplitudes across brain regions. There was also a significant main effect of Condition, ($F(2.62, 70.61) = 4.00$, $p = .014$), suggesting that amplitudes varied across stimulus conditions. There were no significant interactions, indicating that the effect of Condition did not significantly differ across ROIs. All comparisons produced significant projected amplitude responses across all three ROIs (see Table 2). At the first harmonic, amplitudes were largest in the occipital ROI, with the *none vs intact* comparison eliciting the strongest response across all ROIs. The fully-matched and kurtosis-matched comparisons produced intermediate responses, while the variance-added comparison showed the weakest amplitude among conditions—though still significantly above the noise. A similar pattern was observed in the left and right temporal ROIs (see Figure 11).

Third Harmonic

The ANOVA at the third harmonic, with Greenhouse-Geisser corrections applied due to violations of sphericity, revealed a significant main effect of ROI, ($F(1.29, 34.87) = 7.77$, $p = .005$), with no other main effects or interactions. Similar to the first harmonic, all comparisons at the third harmonic produced significant projected amplitudes (see Table 2), although the responses

were lower in amplitude and more uniform across ROIs and conditions compared to the first harmonic, with no major differences between comparison types (see Figure 11).

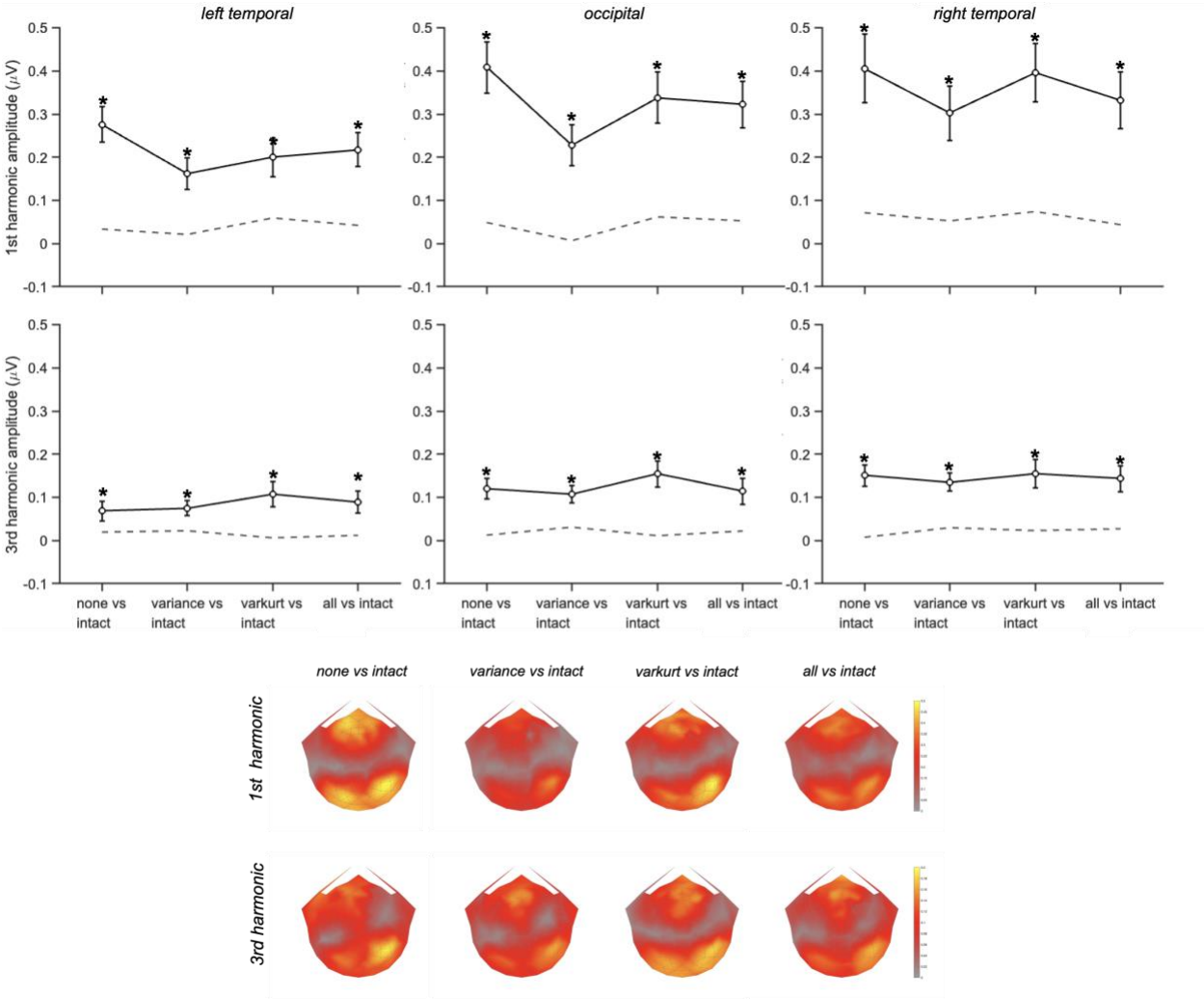

Figure 11: Study 2 Coherent-Averaged Amplitudes Over 3 Electrode Regions of Interest and Topographies. Figure follows the same principles as Figure 9, dotted line is the noise floor, error bars indicate the standard error of the mean. Topographic maps display distributed brain responses for each condition.

	df = 2,27	left temporal		occipital		right temporal	
	condition	*p* value	T^2-statistic	*p* value	T^2-statistic	*p* value	T^2-statistic
1st harmonic	none vs intact	< 0.0001	44.5146	< 0.0001	46.9876	< 0.0001	28.2058
	variance vs intact	< 0.0001	19.7552	0.0002	24.8101	< 0.0001	30.6795
	varkurt vs intact	< 0.0001	19.821	< 0.0001	32.0892	< 0.0001	34.4795
	all vs intact	< 0.0001	32.5538	< 0.0001	36.6974	< 0.0001	26.8706
3rd harmonic	none vs intact	0.0162	10.0801	0.0002	24.4662	< 0.0001	38.9742
	variance vs intact	0.0012	18.4011	< 0.0001	29.2725	< 0.0001	41.0318
	varkurt vs intact	0.0047	13.7955	< 0.0001	35.2059	< 0.0001	26.7501
	all vs intact	0.0085	11.9482	0.0031	15.0593	< 0.0001	21.7122

Table 2: Study 2 Hotelling's T^2 Tests
Table 2 follows the same principles as Table 1. Hotelling's T^2 tests compared each stimulus condition to the noise floor. The tests report the test statistics, degrees of freedom, and p-values. Significant differences reflect robust neural responses to the manipulated shape features.

3.4 Study 3

First Harmonic

In Study 3, we examined the effects of configural disruption on object processing using a 2 × 2 design, comparing Frankenstein with intact animals for both strong and weak Frankenstein, and for upright and inverted version of the stimuli. For the first harmonic, a three-way repeated-measures ANOVA was conducted with factors ROI (three levels), Orientation (two levels), and Strength (two levels), with Greenhouse-Geisser corrections applied due to violations of sphericity. There were no significant effects or interactions (ROI: $F(1.75, 50.74) = 1.63$, $p = .208$; Orientation × ROI: $F(1.40, 40.64) = 0.62$, $p = .486$; Strength × ROI: $F(1.56, 45.16) = 0.09$, $p = .865$; Orientation × Strength × ROI: $F(1.69, 48.99) = 0.19$, $p = .792$). Despite the absence of significant main effects or interactions, we nevertheless ran Hotelling's T^2 tests to test whether any condition

produced a significant response. This exploratory analysis revealed a significant response observed only in the upright strong *Frankenstein vs intact upright* comparison, and only at the first harmonic in the right temporal ROI. No other comparisons, whether involving weak Frankensteins or inverted conditions, produced responses that rose above the noise floor across any ROI (see Figure 12 and Table 3).

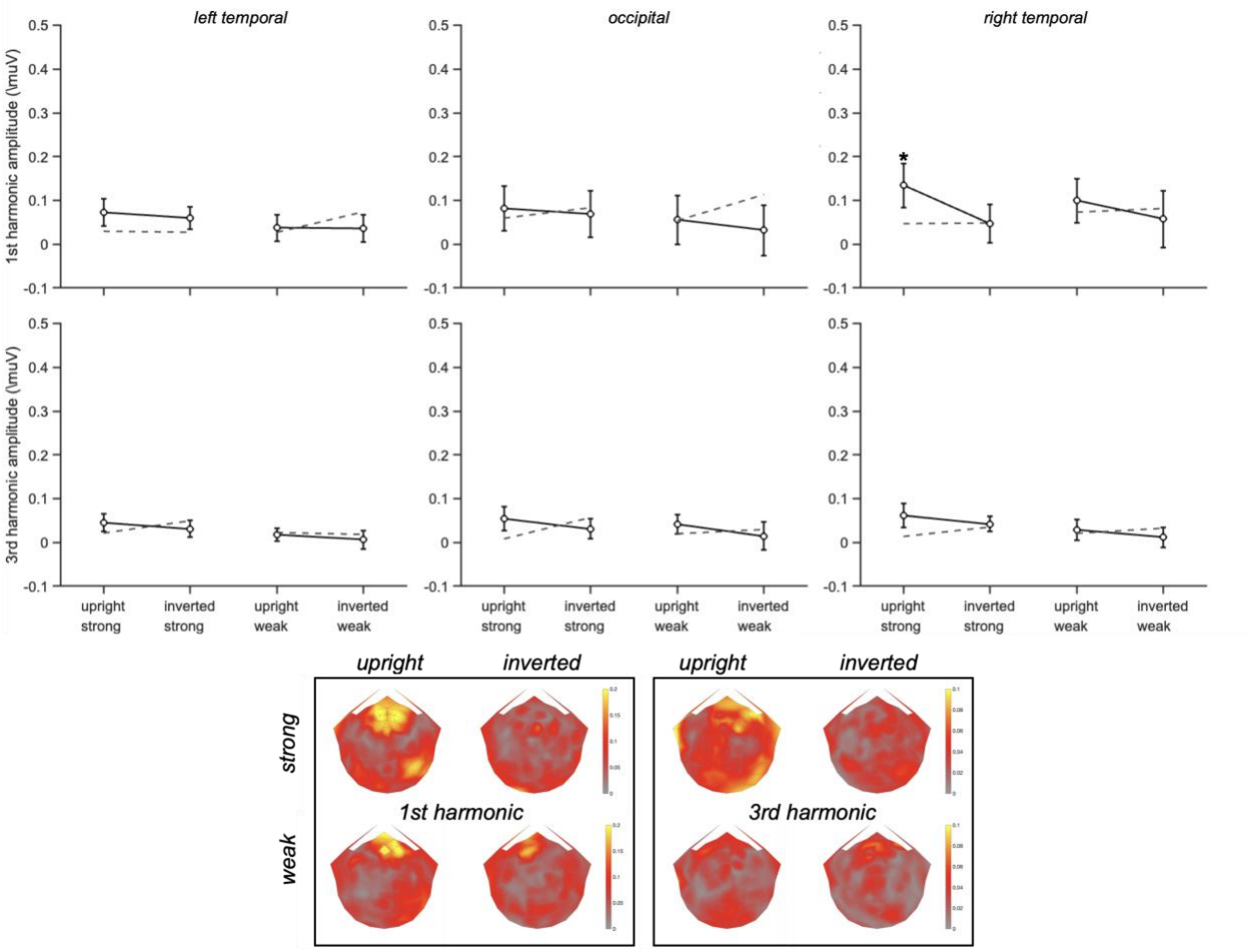

Figure 12: Study 3 Coherent-Averaged Amplitudes Over 3 Electrode Regions of Interest and Topographies. Figure follows the same principles as Figure 9, dotted line is the noise floor, error bars indicate the standard error of the mean. Topographic maps display distributed brain responses for each condition.

Third Harmonic

For the third harmonic in Study 3, the three-way repeated-measures ANOVA with Greenhouse-Geisser corrections applied due to violations of sphericity, found no significant main

effects or interactions (ROI: $F(1.76, 51.08) = 0.39$, $p = .652$; Orientation × ROI: $F(1.60, 46.35) = 0.16$, $p = .807$; Strength × ROI: $F(1.72, 49.86) = 0.35$, $p = .677$; Orientation × Strength × ROI: $F(1.56, 45.16) = 0.01$, $p = .971$). The Hotelling's T^2 tests revealed that no condition produced a response that was significantly different from the noise across any ROIs (see Figure 12 and Table 3).

df = 2,29 condition			left temporal		occipital		right temporal	
			p value	T^2-statistic	*p* value	T^2-statistic	*p* value	T^2-statistic
1st harmonic	strong	frank vs intact upright	0.0885	5.4885	0.3046	2.5701	0.0463	7.1162
		frank vs intact inverted	0.0685	6.1206	0.4244	1.831	0.5857	1.1296
	weak	frank vs intact upright	0.4576	1.6655	0.6203	1.0064	0.1566	4.1068
		frank vs intact inverted	0.43	1.8019	0.7828	0.5116	0.6799	0.8104
3rd harmonic	strong	frank vs intact upright	0.1071	5.0178	0.1638	4.001	0.1072	5.0139
		frank vs intact inverted	0.261	2.92	0.4264	1.8207	0.0681	6.1346
	weak	frank vs intact upright	0.4973	1.4838	0.1632	4.0093	0.5091	1.4326
		frank vs intact inverted	0.9542	0.0973	0.8992	0.2209	0.8708	0.288

Table 3: Study 3 Hotelling's T^2 Tests
Table 3 follows the same principles as Table 1. Hotelling's T^2 tests compared each stimulus condition to the noise floor. The table reports the test statistics, degrees of freedom, and p-values. Significant differences reflect robust neural responses to the manipulated shape features.

4 Discussion

Across our three studies, we found evidence that both local and configural shape information elicit measurable brain responses. Specifically, we observed reliable neural signals reflecting sensitivity to differences in local curvature statistics, disruptions of configural shape, and the effects of stimulus inversion. These results confirm that multiple aspects of shape information contribute to how the visual system represents objects, with distinct patterns emerging

across different experimental manipulations and frequency components of the SSVEP signal. Taken together, our findings highlight the complexity of shape perception and suggest that both local features and configural relationships contribute to how shapes are encoded in the brain. In the sections that follow, we discuss evidence for local shape processing, configural shape processing, and the influence of inversion, before considering how these results inform our understanding of the neural mechanisms underlying shape perception.

4.1 Evidence of Local Shape Processing

Our results provide strong support that the visual system is sensitive to local shape information, particularly reflected in the third harmonic responses. In Study 1, significant third harmonic activity emerged for the *none* vs *variance* comparison especially in the occipital ROI, although the response is more uniform across all three ROIs. This aligns with previous research that showed early visual areas can detect changes in local contour variability (Norcia et al., 2015). In Study 2, while all comparisons of CMC shapes with intact animal silhouettes produced significant responses, the data reveal that sensitivity to local shape appears strongest for variance-related changes. There was a sharp drop in response amplitude from the *none* vs. *intact* comparison to the *variance* vs. *intact* comparison, and the amplitudes remained lower and showed only minor changes with the addition of kurtosis and higher-order moments. This pattern suggests that variance in local curvature is the key feature providing any added value to shape processing, while further statistical refinements, such as kurtosis and higher-order correlations, are too subtle to elicit strong differential neural responses at this stage. Together, these findings indicate that while local curvature features are sufficient to drive distinct neural responses, the visual system may prioritize broader variability signals over finer statistical details when representing local shape. Behavioral

evidence also points to the perceptual relevance of higher-order curvature statistics. Fründ and Elder (2013) showed that observers are sensitive to low-order moments of curvature distributions, such as variance, as well as higher-order moments, and that these cues contribute to shape discrimination. This suggests that the neural sensitivity to higher-order curvature statistics observed here, although modest, may still be consequential for behaviour.

4.2 Evidence of Configural Shape Processing

Our findings also provide clear evidence that the brain is sensitive to configural shape information, particularly reflected in first harmonic responses. In Study 1, the *all vs intact upright* comparison consistently elicited the strongest first harmonic amplitudes, especially in the right temporal ROI. This pattern suggests engagement of higher-level visual areas associated with encoding configural shape (Rossion & Gauthier, 2002; Maurer et al., 2002; Behrmann & Plaut, 2013). In Study 3, we further tested configural shape processing using Frankenstein stimuli, which disrupt the spatial relationships between parts of a silhouette while preserving local shape details. The responses were lower in amplitude relative to Study 1, likely reflecting the more subtle nature of the manipulations: Frankenstein shapes are still animal-like and share many local features with the intact silhouettes, which likely reduces the neural contrast. The Hotelling's T-tests revealed that the first harmonic remains a sensitive index of neural responses to disruptions in configural shape, with significant responses emerging for the *Frankenstein vs intact upright* comparison in the right temporal ROI, but only for the stimuli where the Frankenstein manipulation had a strong behavioural effect. This finding indicates that the first harmonic signal is not merely responsive to changes in configural shape but specifically reflects neural sensitivity to meaningful disruptions in configural relationships that affect object recognition. Significant responses emerge only when

the Frankenstein manipulation produces a strong behavioural effect, and when the stimuli are presented upright. This suggests that these responses provide a direct readout of the neural mechanism sensitive to configural shape that drives the behavioural results found by Baker and Elder (2022). However, it is important to acknowledge that all stimuli in the present study depicted animate objects. Given that animacy is an organizing dimension in ventral visual cortex (Bao et al., 2020; Konkle & Caramazza, 2013), some of the observed effects may have been modulated by category. Moreover, intact animal silhouettes are not only configural but also ecologically familiar, raising the possibility that familiarity could have contributed to the observed effects. Future work should examine whether similar neural sensitivities are observed with inanimate stimuli as well as find a way to control for familiarity to establish the generality of these effects across object domains. Together, our findings support the view that the first harmonic captures sustained neural processes specialized for configural shape representations, consistent with prior work highlighting right-lateralized temporal regions in configural shape processing (Rossion, 2008; Yin, 1969).

4.3 Inversion Effects

Our results also clarify how inversion impacts neural responses to shape. In Study 1, we observed a pronounced reduction in first harmonic amplitude for inverted stimuli compared to upright ones, particularly in the right temporal ROI. This suggests that inversion disrupts the sustained neural mechanisms involved in processing configural shape, consistent with prior evidence that inversion interferes with configural shape representations (Yin, 1969; Rossion, 2008). In contrast, responses at the third harmonic remained significant under inversion, albeit at reduced amplitude. This persistence indicates that although inversion disrupts configural shape processing, it does not abolish it entirely: The visual system retains some capacity to differentiate

configural differences even when stimuli are inverted, albeit with weaker responses. The significant drop in first harmonic responses under inversion supports the notion that this harmonic reflects configural processing mechanisms that are strongly orientation-sensitive. In contrast, the persistence of third harmonic responses under inversion suggests that early visual mechanisms continue to detect differences in contour variability and configural shape even without supporting higher-level categorical encoding. These more transient responses may reflect sensitivity to curvature-based differences rather than configural relationships tied to object categorization. It is also worth noting inversion effects may interact with stimulus category and familiarity, as prior work suggests stronger inversion costs for animate and socially relevant stimuli (Stein et al., 2012). Thus, the observed reduction in first harmonic responses for inverted silhouettes may reflect a convergence of configural sensitivity and animacy-related processing.

4.4 Two Mechanisms

Our findings support the existence of two distinct mechanisms in shape perception. The first is a more sustained, higher-level mechanism, indexed by the first harmonic, sensitive to configural relationships among shape parts. This mechanism is strongly impacted by stimulus inversion and localized to right temporal regions. The second is a more transient, early visual mechanism, indexed by the third harmonic, that responds to variations in statistics of local shape, specifically variance, as well as configural shape. This mechanism appears less sensitive to inversion and is primarily centered in occipital cortex. Importantly, while the specific timing of the harmonics depend on the fundamental frequency, lower harmonics (e.g., 1f1) are inherently more sustained than higher harmonics (e.g., 3f1), In the present experiments, with a base frequency of 1 Hz, first harmonic reflects relatively slow, sustained responses, while the third harmonic reflects relatively faster, transient responses. This interpretation aligns with prior reports linking

lower harmonics of slow stimulation rates to sustained, higher-level processes, and higher harmonics to transient, early visual responses (Norcia et al., 2015).

Evidence for these two mechanisms emerges clearly across our studies. The first harmonic was most robust in conditions involving intact animal silhouettes and in disruptions to configural shape, such as the Frankenstein stimuli, suggesting engagement of higher-level processes sensitive to configural shape. These responses were strongest in right temporal regions, consistent with prior research linking this area to configural object perception (Rossion & Gauthier, 2002; Maurer et al., 2002). Inversion effects further highlight the role of this mechanism, as first harmonic responses were markedly reduced when stimuli were inverted, indicating a strong orientation sensitivity characteristic of configural processing.

In contrast, the third harmonic responses were prominent for local shape manipulations, particularly evident in the *none* vs *variance* comparison. These results indicate that early visual areas, reflected in occipital activity, are tuned to differences in local shape properties, aligning with previous work showing higher harmonics as signatures of early, transient visual processing (Ales et al., 2012; Liu-Shuang et al., 2014). Interestingly, while local shape differences were detectable, Study 2 showed that finer adjustments beyond variance, such as kurtosis and higher-order moments, led to diminishing neural differentiation, suggesting that the third harmonic is especially sensitive to broader contour variations. Future research should extend these findings by systematically varying animacy and familiarity, for instance by directly comparing animate and inanimate silhouettes matched for low-level features. Such work will clarify whether the harmonic responses reported here reflect configural processing rather than familiarity.

Together, these findings suggest that shape perception in the visual system emerges from an interplay between these two mechanisms: one specialized for configural integration and another

for local shape. This dual-process framework provides a neural basis for how the brain combines local details into coherent configural representations, enabling object recognition.

References

Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *Journal of Vision, 12(*10), 18. https://doi.org/10.1167/12.10.18

Baker, N., Lu, H., Erlikhman, G., and Kellman, P.J. (2018). Deep convolutional networks do not classify based on global object shape. PLoS. *Computational Biology. 14*, e1006613.

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2020). Deep convolutional neural networks do not classify based on global object shape. *Vision Research, 172*, 46–61. https://doi.org/10.1016/j.visres.2020.04.003

Baker, & Elder, J. H. (2022). Deep learning models fail to capture the configural nature of human shape perception. *iScience, 25*(9), 104913–104913. https://doi.org/10.1016/j.isci.2022.104913

Bao, P., She, L., McGill, M., & Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. Nature, 583(7814), 103–108. https://doi.org/10.1038/s41586-020-2350-5

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences, 17(*5), 210–219. https://doi.org/10.1016/j.tics.2013.03.007

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Biederman, I., and Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*, 38–64.

Blumenfeld, H. (2022). Neuroanatomy through clinical cases (3rd ed.). Sinauer Associates.

Brandman, T., & Yovel, G. (2012). A face inversion effect without a face. *Cognition, 125*(3), 365-372.

Brincat, S. L., & Connor, C. E. (2007). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 10(*2), 205–213. https://doi.org/10.1038/nn1836

Bushnell, B. N., Harding, P. J., Kosai, Y., & Pasupathy, A. (2013). Equiluminance cells in visual cortical area V4. *Journal of Neuroscience, 33*(40), 16383–16395. https://doi.org/10.1523/JNEUROSCI.2147-13.2013

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS *Computational Biology, 10*(12), e1003963. https://doi.org/10.1371/journal.pcbi.1003963

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex, 17*(3), 713–731. https://doi.org/10.1093/cercor/bhk023

Carey, S. (1992). Becoming a face expert. Philosophical Transactions of the Royal Society of London B: *Biological Sciences, 335*, 95–103.

Cottereau, B. R., Ales, J. M., & Norcia, A. M. (2011). Increasing the accuracy of electromagnetic inverse solutions using functional area source correlation constraints. *Human Brain Mapping, 32*(9), 1532–1547. https://doi.org/10.1002/hbm.21125

Elder, J. H., Oleskiw, T. & Fründ, I (2018). The Role of Global Cues in the Perceptual Grouping of Natural Shapes. *Journal of Vision 18*, 1–21.

Elder, J. H., & Velisavljević, L. (2009). Cue dynamics underlying rapid detection of animals in natural scenes. *Journal of Vision, 9*(7), 1–16. https://doi.org/10.1167/9.7.7

Fründ, I., & Elder, J. H. (2013). Statistical coding of natural closed contours. Journal of Vision, 13(9):119. doi:10.1167/13.9.119

Gauthier, I., Tanaka, J. W., & Brown, D. D. (2003). When misaligned faces are processed holistically. *Journal of Vision, 3*(9), 92. https://doi.org/10.1167/3.9.92

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., and Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. https://doi.org/10.48550/arXiv.1811.12231.

Gerlach, C. (2025). Large inversion effects for common objects and objects that require little if any holistic processing. *Scientific Reports, 15*(1), 7275. https://doi.org/10.1038/s41598-025-92035-3

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41(*10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hegde, J., & Van Essen, D. C. (2001). Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience, 21*(5), 1255–1266. https://doi.org/10.1523/JNEUROSCI.21-05-01255.2001

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 198*(1130), 1–59. https://doi.org/10.1098/rspb.1977.0085

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*(4), 289–303. https://doi.org/10.3758/BF03197689

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology, 10*(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Kohler, PJ, Clarke, A, Yakovleva, A, Liu, Y & Norcia, AM (2016). Representation of maximally regular textures in human visual cortex. *Journal of Neuroscience, 36*(3), 714 –729.

Konkle, T., & Caramazza, A. (2013). Tripartite Organization of the Ventral Stream by Animacy and Object Size. Journal of Neuroscience, 33(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. Science, 293(5534), 1506–1509. https://doi.org/10.1126/science.1061133

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. Journal of Experimental Psychology: Human Perception and Performance, 22(2), 395–416. https://doi.org/10.1037/0096-1523.22.2.395

Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia, 52*, 57–72. https://doi.org/10.1016/j.neuropsychologia.2013.10.022

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences, 6*(6), 255-260.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review, 16*(5), 778–797. https://doi.org/10.3758/PBR.16.5.778

Norcia, Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. J*ournal of Vision (Charlottesville, Va.), 15*(6), 4–4. https://doi.org/10.1167/15.6.4

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), Attention and performance IX (pp. 135–151). Erlbaum.

Palmer, S. E. (1999). Vision science: Photons to phenomenology. MIT Press.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology, 82*(5), 2490–2502. https://doi.org/10.1152/jn.1999.82.5.2490

Pomerantz, J. R. (1986). Visual form perception: An overview. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and human performance (Vol. II, pp. 1–46). Wiley.

Prasad, S., & Galetta, S. L. (2011). Anatomy and physiology of the afferent visual system. *Handbook of Clinical Neurology, 102*, 3–19. https://doi.org/10.1016/B978-0-444-52903-9.00001-3

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron, 74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Rossion B, Gauthier I. (2002). How does the brain process upright and inverted faces? Behav Cogn Neurosci Rev. Mar;1(1):63-75. doi: 10.1177/1534582302001001004. PMID: 17715586.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. Acta *Psychologica, 128*(2), 274–289. https://doi.org/10.1016/j.actpsy.2008.02.003

Rossion, B., & Curran, T. (2010). Visual expertise with pictures of cars correlates with RT magnitude of the car inversion effect. *Perception, 39*(2), 173–183. https://doi.org/10.1068/p6151

Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *Journal of Neuroscience, 30*(39), 12978–12995. https://doi.org/10.1523/JNEUROSCI.0179-10.2010

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: Evidence from inversion effects during continuous flash suppression. Cognition, 125(1), 64–79. https://doi.org/10.1016/j.cognition.2012.06.005

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46*(2), 225–245. https://doi.org/10.1080/14640749308401045

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). MIT Press.

Willmore, B. D. B., Prenger, R. J., & Gallant, J. L. (2010). Neural representation of natural image structure in primary visual cortex. *Journal of Neurophysiology, 104*(4), 2103–2116. https://doi.org/10.1152/jn.01013.2009

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences, 111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology 81*, 141.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.