

2019

What causes the dip in object recognition rotation functions?

Charles Josef Peasley
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>



Part of the [Cognitive Psychology Commons](#)

Recommended Citation

Peasley, Charles Josef, "What causes the dip in object recognition rotation functions?" (2019). *Graduate Theses and Dissertations*. 17538.
<https://lib.dr.iastate.edu/etd/17538>

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

What causes the dip in object recognition rotation functions?

by

Charles Josef Peasley

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Major: Psychology

Program of Study Committee:
Eric Cooper, Major Professor
Jonathan Kelly
Kevin Blankenship

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Charles Josef Peasley, 2019. All rights reserved.

TABLE OF CONTENTS

	Page
LIST OF FIGURES.....	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
CHAPTER 1. GENERAL INTRODUCTION.....	1
Background	1
Structural Description Theories	3
Template Theories	11
The “dip” in the recognition time function at 180°	15
CHAPTER 2. EXPERIMENT ONE.....	19
Methods	22
Results	24
Discussion	27
CHAPTER 3 EXPERIMENT TWO.....	28
Methods	31
Results	35
Discussion.....	37
CHAPTER 4 GENERAL DISCUSSION	39
REFERENCES	45

LIST OF FIGURES

		Page
Figure 1	A picture of the 36 geons posited by RBC	6
Figure 2	Line drawings of a coffee mug and a bucket	9
Figure 3	Three coffee mugs with the same structural description	10
Figure 4	Coffee mug under planar rotation	17
Figure 5	A side of and an above-below object rotated	20
Figure 6	Predicted rotation functions for above/below and side-of objects	21
Figure 7	Experiment 1 Stimuli	23
Figure 8	Experiment 1 Reaction Time Results	25
Figure 9	Experiment 1 Error Rate Results	26
Figure 10	Above-below and side-of object with consistent geometry	29
Figure 11	Experiment 2 distractor objects with part identity swapped	32
Figure 12	Experiment 2 distractor objects with part positions swapped	32
Figure 13	Experiment 2 procedure for same and different trials	33
Figure 14	Experiment 2 Reaction Time Results	36
Figure 15	Experiment 2 Error Rate Results	37

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Eric Cooper, and my committee members, Dr. Kevin Blankenship, and Dr. Jonathan Kelly.

Additionally, I would like to thank Dr. Alex O'Brien and Dr. Jeremiah Still for their input at conferences which influenced this work at various stages, and Alexander R. Toftness for his help brainstorming objects that were used in the first experiment.

ABSTRACT

Two experiments were conducted to determine why there is a local improvement in recognition times when object images are inverted. Experiment 1 used naturally occurring, everyday objects and measured the effects of picture plane rotation on their identification times. Performance varied according to spatial configuration type, wherein only side-of objects become easier to recognize upon complete inversion than at neighboring orientations, forming a “dip”. Above-below objects became increasingly difficult to recognize as rotation approaches 180 degrees. Experiment 2 employed novel non-sense objects in a sequential matching paradigm. Rotation function shapes displayed an interaction in the same direction as Experiment 1, though no “dip” in response times was observed. In Experiment 2, experimenter prescribed categorical part relations influenced the shape of rotation functions for recognition independent from other object properties. Rotation functions revealed that obtaining this counterintuitive local improvement depends upon the presence of side-of relations between an object’s parts. Together, these experiments provide evidence for the use of categorically coded structural descriptions in object recognition.

CHAPTER 1. GENERAL INTRODUCTION

Background

The field of artificial intelligence has made monumental achievements in recent history by creating computerized systems which are able to match human performance on a myriad of difficult tasks fueled by advancements in processing power and access to large training datasets. Perhaps most famously, AlphaGo beat grandmasters in chess and the game of Go, which is a very complex competitive challenge requiring the consideration of thousands of moves per turn (Yost, 2012). In addition, artificial intelligences such as CAN demonstrate computers can create visual art that is effectively indistinguishable from the products of human creativity (Elgammal et al, 2017). Despite these, among many other fantastic achievements, a generic object recognition computer vision protocol remains elusive. Certain tasks like face recognition have essentially been solved as common consumer products boast effective identity verification by using a person's face as a biometric input. However, computer vision algorithms generally perform recognition tasks at a much lower success rate than the average human (Ullman et al, 2016). Minor changes in the way an image is presented can have drastic negative impact on identification accuracy which suggests a need for representations that are resilient to changes in image viewpoint (D'Innocente et al, 2017; Ghodrati et al., 2014).

One potential framework for developing programs that match or even outperform humans involves designing algorithms based on the way humans perceive objects. Supporting the validity of this proposal, Schrimpf et al. (2018) charted the "brain-scores" of deep artificial neural networks, or the degree to which their internal representations resemble the brain's mechanisms for core object recognition, and found that the more similar the neural networks' operations are to the human visual system, the better they perform on the computer vision

benchmark testing dataset, ImageNet. The authors call for a deeper understanding of human pattern recognition to guide and simplify the development of ever more efficient machine networks. Many other researchers have stressed the importance of using the human visual system to inform the development of artificial vision systems (González-Casillas et al., 2018; He, Yang, & Tsien, 2011; Kruger, Lappe, & Worgotter, 2003; Kubilius et al., 2018; Sinha et al., 2006)

Although taken for granted in daily life, the ability to recognize visual stimuli subserves a large portion of our interactions with the environment. Determining object identity enables a plethora of higher-order cognitive procedures such as appropriate judgment and decision making, tool use, and categorization (Hummel, 2013). Despite many years of research conducted across numerous laboratories, contention exists between theories attempting to explain how the human visual system solves object recognition (Gauthier & Tarr, 2016). The field must resolve these debates before the creation of analogously functioning artificial systems is possible with any degree of certainty. Beyond the technological applications, the etiology of visual recognition disorders like dyslexia and prosopagnosia, or face-blindness, remain mysterious. Determining how feats of visual recognition are typically accomplished may offer critical insights for designing treatment programs or even reveal opportunities for preventative intervention. Toward these broad motivations, the research described in this paper compares contradicting predictions made by two prominent classes of object recognition approaches for the effects of rotation on identification in two distinct experimental paradigms. The immediately following sections will describe these two theoretical approaches in more detail with the motivation of ascertaining the unique predictions that follow from each.

Structural Description Theories

A major challenge for any theory of object recognition involves reconciling the inverse optics problem. The retina receives information in a 2-D plane, and yet we experience the perception of depth in the environment. The inverse optics problem acknowledges that any 2-D image could be created by an infinite number of 3-D environments due to variations in viewing distances, orientation, and pose. To compensate for this uncertainty, our visual system must prefer reliable signals that tend to convey trustworthy information about the nature and form of our visual environment across time. Lowe (1985) proposed a set of visual elements which could subserve invariant recognition and solve the inverse optics problem to a reasonable extent. These elements, coined non-accidental properties (NAPs), are components of shape that are extremely robust to changes in viewpoint. When an object is transformed or translated in space, NAPs are generally maintained across object instances, including whether edges are straight, convex, or concave. Other examples of these stable properties include parallelism, collinearity, equal spacing, and symmetry (see Lowe 1985 for a more detailed description). NAPs are distinct from metric properties (MPs) which involve the specific size, position, and orientation of image features. For examples of each type of property, imagine two straight lines that run parallel to each other with a separation of two arbitrary units. Here, the NAPs are the descriptors straight and parallel, while the MP mentioned is the precise distance between the lines (2 arbitrary units). These two lines will never become perpendicular or orthogonal as an observer is exposed to different viewpoints and neither will they become curved. However, depending on the distance in depth that their image is presented from an observer, the metric distance between them may appear to deviate from two units.

Only under occlusion or at infrequent chance occurrences, called *accidental angles*, are NAPs indiscernible from an image or inconsistent with the true three-dimensional reality of an object. If a two-dimensional drawing of an object contains a curved line the visual system infers the corresponding presence of a curved feature in three-dimensions. In the rare instance of an accidental alignment the curved line could fall on the retina in such a way as to appear straight. These instances are extremely unlikely. Due to their reliability, NAPs are hypothesized to be more important for object recognition than MPs which vary drastically over alterations in perspective. For example, the degree of curvature a rounded line projects varies continuously as its three-dimensional image is rotated in depth, rendering this metric information relatively undiagnostic and inefficient to rely on for robust identification. Amir, Biederman, & Hayworth (2012) found that changes to NAPs of geons were more noticeable than changes to their MPs even when the changes to geon MPs were slightly greater according to pixel space.

Computer models designed to differentiate between basic geometric images show support for the notion that NAPs relay particularly diagnostic information about shape. Neural networks using a biologically plausible learning rule, called temporal tracing, demonstrate preferential encoding for NAPs in simple objects as compared to metric properties like the length of individual lines (Rolls & Mills, 2018). The unsupervised learning network in this study developed a hierarchical encoding scheme in which each progressive layer became more generalized or view-independent without explicit direction to organize in that fashion. The primary layer responded to lines and their orientation, intermediate layers responded to general features and feature combinations contained in more than one object type, and the final layer was remarkably object selective and invariant to translations in stimuli presentation.

In striking correspondence, the organization of the human ventral inferotemporal cortex, commonly understood to be the center of invariant visual recognition, supports an analogous hierarchical coding scheme (Rolls & Tovee, 2017; Yamins et al., 2014). Neuronal layers have progressively larger receptive fields or regions of visual space for which stimuli elicit a response. The primary visual cortex contains a retinopic map that retains view specific details, while the inferotemporal cortex ultimately responds to objects independent of most viewing conditions (Hubel & Weisel, 1974; Tanaka & Taylor, 1991). As the neural network model of Rolls and Mills (2018) was not supervised or directly instructed to attune to any given feature other than those that are most useful for discriminating between images, the use of NAPs is also theoretically consistent with modern Bayesian theories of human perception which assert human visual perception is based upon calculating statistical probability distributions and expectations inspired by experience (Rolls & Mills, 2018; Yuille & Kersten, 2006).

All object recognition approaches describe recognition as a process by which an impinging image is compared to representations stored in memory. At the time of Lowe's seminal paper describing NAPs, most accounts of human object recognition posited the use of volumetric primitives or three-dimensional shapes in the construction of object memory representations (Marr, 1982; Lowe & Binford, 1983). *Structural description theories* such as Marr's posit that visual identity is determined by encoding the individual parts making up an object and describing the structural relationship amongst these parts to ultimately find the closest match to this description stored in memory. These theories explain that objects are partitioned into separate groups of features which contain Gestalt-like visual similarities, together forming distinct volumetric parts. This feature-grouping is alternative to every visual feature being represented holistically, as discussed in the competing visual theories of a subsequent section. In

structural description theories, the accuracy and time course for recognizing an imaged object depends on the extent to which depicted parts and the relationship between them matches that of a representation in memory. Neural evidence using macaque monkeys as model organisms demonstrated that the IT (responsible for invariant recognition) is tuned for detecting the three-dimensional spatial configuration of simple three-dimensional primitives (Yamane et. al, 2008).

One of the most well-known and cited object recognition theories is Recognition-by-Components (RBC), which combines aspects of original structural description theories with the viewpoint robustness of NAPs. Biederman (1987) expanded upon contemporaneous structural description theories by proposing a specific set of visual primitives called geons, short for “geometric icons”, based on all computationally possible combinations of NAPs ($N \leq 36$; see Figure 1 for an image array of each geon).

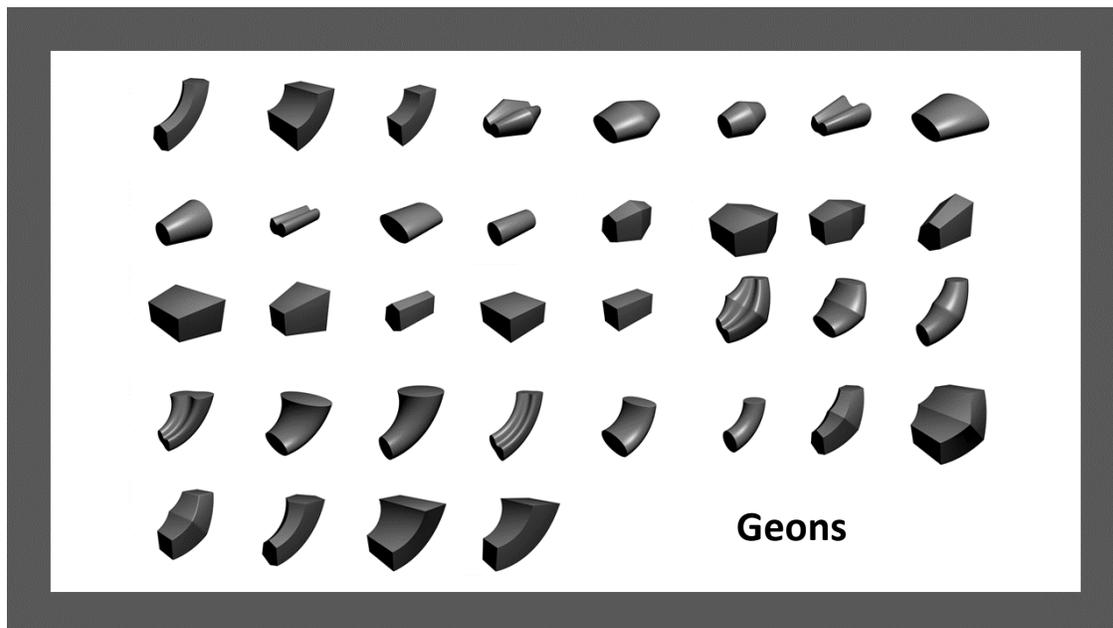


Figure 1. Picture of the 36 geons posited by RBC (courtesy of O’Brien, 2018). Renderings were created using Blender 2.79.

For an example, a brick-shaped geon is defined by its straight edges, a constant sized cross section, rotational and reflective symmetry, and a straight axis. As such, a straight cylinder geon differs from the brick geon in having curved edges instead of straight edges. Just as the relatively small set of phonemes in the English language combines to form tens of thousands of words, geons theoretically provide basic building blocks with the representational power to compose the tens of thousands of everyday objects we encounter. In general, part relations entail comparisons made between visual primitives in their respective size, orientation, and spatial configuration or position, thus offering greater degrees of freedom than the temporal restrictions of spoken language.

In addition to defining a set of primitives, RBC proposes that the part relations between geons (size, orientation, and positions) are encoded categorically. Categorical perception occurs when a continuously varying stimulus is perceived in discrete categories. For example, though the electromagnetic spectrum varies continuously, the human visual system classifies specific ranges of visible light into discrete color categories like blue or red (Harnad, 1987). Categorical perception allows a receiver to overcome the problem of infinite environmental variation and extract discrete commonalities and differences among stimuli. In RBC, the categorical perception of size proceeds by classifying geons as relatively smaller than, larger than, or equal in size. In a drawing task, non-expert artists demonstrated heavy reliance upon categorical size relations when recreating images with multiple shapes (Arnold, 2014). Orientation would be encoded via the categories parallel, perpendicular, and oblique. Rosielle and Cooper (2001) discovered that objects were difficult to discriminate when a change in the orientation of their parts fell within the same general category (i.e., oblique vs. more oblique) supporting the notion that orientation is coded categorically. Lastly, RBC posits that the categories used for encoding a

visual primitive's spatial relations are above, below, and side of. Kranjec, and colleagues (2014) found robust perceptual biases for classifying spatial relations in categorical terms using dot-cross configurations. Dot positions were more readily described and differentiated when appearing above, below, or to the side of a large X-shaped cross than when two dots appeared in the same visual category. This effect held even when holding the magnitude of the metric difference between the dot positions constant across conditions. Additionally, the relations above, below, and side-of were preferred to combinations therein (e.g. both above and to the side-of).

Crucially, incorporating part relations into structural descriptions enables visual discrimination between objects which share parts. Continuing the analogy with language, the serial ordering of phonemes enables us to distinguish between words composed of the same subset of sounds. For example, "rough" and "fur" contain the same phonemes but are perceived as different words. By analogous operations, the perception of spatial relations allows a viewer to tell objects apart when they are composed by same subset of geons. For example, though a coffee mug and a bucket share individual parts in common, we can readily discriminate them by noticing the spatial relationship between these parts differs categorically (see Fig. 2 below). In visual search tasks, objects that share a spatial configuration in common with the target item interfere with search performance because of their perceived similarity, and this effect occurs independently of whether they share parts (Arguin & Saumier, 2004). Wasserman and colleagues (1993) showed that even pigeons are effective at telling complex objects apart based on differences in the spatial configuration of parts.

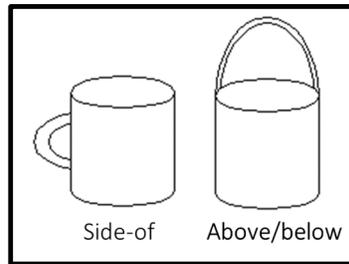


Figure 2. Line drawings of a coffee mug and a bucket. Both objects are composed by a curved cylinder and a straight cylinder. They are differentiated by the relationship amongst their parts.

Notice the distinction in RBC between spatial comparisons made in the lateral and vertical planes. The interchangeable specification of left and right permitted by grouping “side of” part relations together enables the continuous utility of a single structural representation as an observer walks around an object and those ephemeral left/right polarities change, while, in contrast, above/below relations remain constant throughout most naturally occurring, translative changes in an observer’s viewpoint. Behavioral measures demonstrate that human recognition is in fact largely invariant to left-right reflections, as would be predicted by RBC, because any object image has the same structural description as its left-right reflection (Biederman & Cooper, 1991; Fiser & Biederman, 2001; Stankiewicz, Hummel, & Cooper, 1998). Conversely, top-bottom reflections result in a marked decreased in recognition performance (Jolicoeur, 1990; Thoma, Davidoff, & Hummel, 1987). Thus, it appears the human brain preferentially distinguishes between above and below and generalizes between left and right relations in basic-level visual representations.

A prediction that follows from this notion is that pattern recognition in which the horizontal order of parts contains essential information, like reading, should result in processing difficulty or require some degree of learning and expertise. Indeed, children and beginning readers require substantial instruction and practice before becoming literate, and people make systematic mistakes while they learn. As RBC theory would predict, Hildreth (1934) and Rice (1930) found that common writing mistakes involved the mirror reflection (often used

synonymously with left-right reflections) of letters dramatically more often than vertical inversions. Children also sometimes pronounced words as if they were read starting right to left instead of left to right. Accordingly, structural description-like representations (called analytical representations here) appear to develop in adolescence prior to the emergence of visual representations that explicitly encode horizontal directionality or handedness (Wakui et al., 2013). These examples support the theoretical coding scheme of spatial relations used for basic level object recognition into the discrete categories posited by RBC theory.

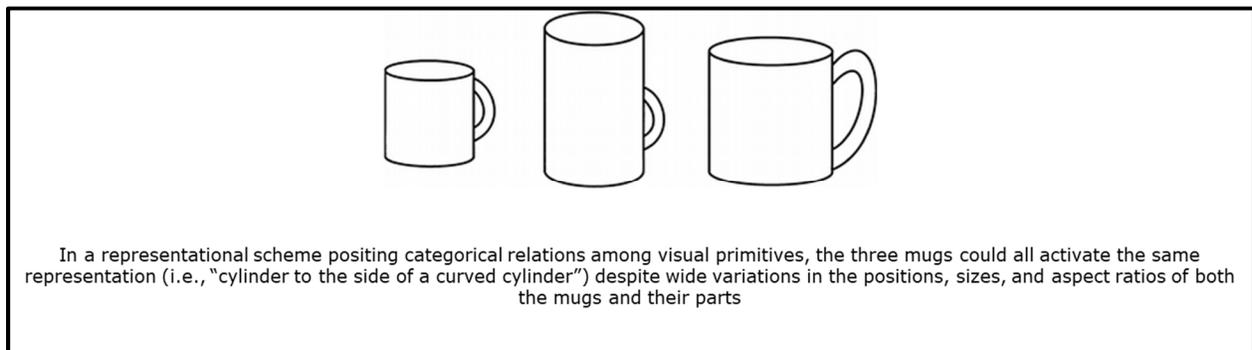


Figure 3. Three coffee mugs with the same structural description. Adapted from Cooper and Wojan (2000)

In RBC, the relations between the volumetric primitives of a coffee mug would be coded in this manner: a curved cylinder (the handle) that is parallel to (orientation), side of (position), and smaller than (size) a straight cylinder (the vessel). Utilizing such broad descriptors enables flexible generalization of object representations to new exemplars of a familiar object set or category. Imagine an individual has never seen a given coffee mug with a such a small handle; despite the unusual Euclidean metric properties of that particular handle, this person would properly recognize the presented object as a coffee mug because structural descriptions tolerate deviations from subordinate representations (Biederman, 1987).

Although structural description theories provide theoretically sound mechanisms for accomplishing invariant object recognition, mirror generalization phenomena, and the

categorization of novel exemplars, behavioral data obtained in planar rotation experiments demonstrates that object recognition performance depends on viewpoint. That is to say, because reaction time increases when objects are rotated in the picture plane (like the hands of a clock), object recognition is not truly invariant to misorientations from a canonical standard view (Gauthier & Tarr, 2016; Peissig & Tarr, 2007). Many investigators likely interpreted Biederman's (1987) assertion that geons are identified with invariance (up to occlusion and rare accidental angles) implies recognition performance should be infallibly robust to manipulations in object orientation and pose that preserve the visibility of NAPs. However, subsequent processing stages described in RBC following the identification of geons, namely the determination of their spatial relations, may very well be impacted by planar rotation. Nevertheless, structural description theories were considerably supplanted by view-dependent representational theories to explain the boundaries of view-invariant object recognition (Edelman & Bülthoff, 1992a; Lawson et. al, 1994; Tarr, 1995). The observation that reaction time functions for naming rotated objects *nearly* mimic reaction times in mental rotation tasks has suggested to some that misoriented objects are identified by mentally rotating their visual image to the standard upright orientation or a learned view (Corballis et. al, 1978; Shepard & Metzler, 1971; Shepard & Cooper, 1982).

Template Theories

View-based recognition theories posit that people store isomorphic copies of exterior patterns observed in the past within long-term memory. Object representations in view-based models are stored as templates: a collection of raw intensity values corresponding to the coordinates of the visible features of the image (Ullman, 1989). Empirical support has accumulated that template-like representations are used for visual tasks requiring differentiation

between objects that share a structural description (e.g. Gauthier et. al, 1999). For example, all human faces have the same parts and categorical relationships between them. Recognizing individual faces likely relies on calculating fine metric differences between parts and their relations, like the size of a nose and the precise location of the mouth (Cooper & Wojan, 2000). Another example of recognition requiring the calculation of precise metric characteristics would be individuating the previously described coffee mug that shares a structural description in common with other coffee mugs but possesses a uniquely small handle. As alluded to earlier, RBC provides no mechanism for individuating members in structurally homogeneous categories, though we are able to tell apart exemplars of a wide variety of stimuli types. Observers readily distinguish faces, coffee mugs, cars, birds, etc. that share structural descriptions. These talents necessitate the operation of a perceptual system with sensitivity to metric variations occurring within categorical boundaries. These cases demonstrate the usefulness of template representations for performing within-category (subordinate) discriminations.

Template models of human object recognition match stimulus input to a single or to multiple two-dimensional internal representations. When presented with a visual pattern, template theories presume the input is compared to all of the two-dimensional internal object representations that have been previously encountered and learned. After normalizing the input to a standard size, rotation, and orientation, identification occurs by selecting whichever representation shares the most intensity values in common with the observed pattern. These models predict increased recognition latency under viewing conditions never previously encountered, such as changes in position, rotation in depth and picture plane (Ullman, 1989; Jolicœur, 1985; Edelman & Bühlhoff, 1992a). View-based theories postulate that shape features are bound to a spatial location in an image and the explicit representation of metric properties.

According to template theories, recognition latency corresponds to the degree of transformation necessary to correct for the mismatch between the stored representation and the input conditions. In template theories, as an object deviates from its canonical viewpoint, an increasingly time-consuming transformation is required to reorient the input image and make a valid comparison to the features and metric properties of stored object representations (Dixon & Just, 1978; Neisser, 1967). For rotation in the picture plane or rotation in depth, the predicted monotonic recognition function implies explicit mental rotation, during which a mental image proceeds in analog through intermediate stages up to normalization while isomorphic to the visual qualities of the image as it would appear when physically rotated in the external world (Cooper & Shepard, 1973; Murray, 1997).

A key characteristic of viewpoint-dependent theories is that multiple representations are stored for a single object (Tarr & Kriegman, 2001). Additional representations of an object's shape are formed after exposure to the object under viewpoints sufficiently different from that of the stored presentation. Forming multiple representations allows for more efficient shape processing across variation in viewing conditions. According to viewpoint-dependent theories, once someone is presented with the image of a palm tree rotated 90° , he or she will recognize the same rotated image of a palm tree faster upon testing by using a new template (Tarr & Pinker, 1989). Latencies for identifying an inverted palm tree (180°) would decrease as well because the mental transformation necessary to rotate that image to 90° is less computationally intensive than rotating the image to the upright orientation (0°) represented by the original template. These predictions have been supported by experiments in which the recognition of rotated objects occurs more expediently after learning them from multiple viewpoints (Jolicoeur, 1985; Koriat & Norman, 1985; Tarr, 1995). With enough exposures and variation in viewpoint, objects are

recognized with equivalent expedition across viewpoints (Tarr & Pinker, 1989). This effect applies to everyday objects that have no standard base or that are effectively poly-oriented. In addition, interpolations between templates acquired from rotations in depth could theoretically allow for recognition of the object from any viewpoint (Edelman and Bulthoff, 1992b; Riesenhuber & Poggio, 2000).

Although template theories account for the presence of identification costs associated with the rotation of an object from upright, there are characteristics of human pattern recognition which these theories inadequately address. Despite template theorists positing a common mechanism for alignment in mental rotation tasks and the recognition of misoriented objects, there are important differences that emerge when the two processes are analyzed in comparison. Mental rotation tasks do not evidence the development of multiple stored views for the objects used in testing. In other words, orientation effects diminish with practice for the stimuli used in object recognition but not for mental rotation experiments (Jolicouer 1985; Tarr & Pinker, 1989). Wilson and Farah (2006) used fMRI to compare the patterns of brain activity evoked by classic mental rotation tasks and misoriented object recognition. The authors found strong evidence that misoriented object recognition is subserved by distinct neural systems from those implicated in performing mental rotation on an image. Although activity in several other brain regions increased compensatory to stimuli's angular disparity from upright, the superior parietal lobe revealed no significant disparities in blood-oxygen-level-dependent activation during the identification of misoriented objects and their upright, and across all levels of rotation. This result was important because the superior parietal lobe is widely considered the biological hub for mental rotation operations and exhibits systematic activity in response to the magnitude of

the corrective orientation transformation necessary for completing a rotation task (Carpenter et al., 1999; see also Gauthier et al., 2002).

There are double dissociations documented in the literature where individuals retain the ability to recognize rotated objects despite drastic impairment in their mental rotation capacities, and vice versa. Morton & Morris (1995) reported a patient who could recognize objects under unusual viewing conditions but performed poorly with forms of mental rotation tests. Farah and Hammond (1988) report a stroke patient who could recognize alphanumeric characters, anagrams, and objects at misorientations and inversion but who could not perform mental rotation. In obverse, Turnbull and McCarthy (1996) reported a patient who could perform mental rotation tasks as well as the controls but was poor at identifying objects when they were rotated in the picture plane. So far, the evidence described suggests that a process in mentally rotating an image to reference a canonical orientation is an unlikely candidate for the underlying operation used when recognizing misoriented objects.

The “dip” in the recognition time function at 180°

An extremely well replicated patterning in picture-plane object recognition time functions is a dip from about 135° to 225° that reaches a local minimum at 180° (Diwadkar & McNamara, 1997; Cooper & Brooks, 2004; Corballis et al., 1978; Fodor et al., 1989; Harris & Dux, 2005; Jolicouer, 1985; Koriat & Norman, 1984, 1985; Large et al., 2003; Murray, 1997; Rock et al., 1981). With an elegantly designed experiment, George Dearborn (1899) provided the first account of superior object recognition upon inversion than what would be expected by interpolating between neighboring orientations. Dearborn speculated that the M-shaped function he obtained reflects natural similarities between an upright object and its visual inverse.

Indeed, structural description theories would explain that this relative improvement manifests because canonical “side of” relationships are restored when an object is rotated 180° (Cooper & Wojan, 2000; Hummel, 1994). Rotation in the picture plane preserves the visibility of an object’s non-accidental properties, and therefore affords unimpeded extraction of part identities, while the nominal spatial relationship between parts may have become perturbed. Structural description theories predict decrements to recognition when an object like a coffee mug has been rotated 90° or 270° because of categorical changes to primitive spatial relationships (see Fig. 4 below). The handle would reside “below” or “above” the curved cylinder in each respective case, leading to a comparative mismatch between the pictured image and the description of a coffee mug in memory (i.e. a curved cylinder “side of” a straight cylinder). Crucially, the handle reestablishes “side of” the main cylinder when its image is presented at 180°. At this rotation the coffee mug’s structure corresponds with its parts and categorical part relations expressed in the stored description, and this agreement facilitates the matching process and increases the likelihood for accurate recognition. As this example demonstrates, structural description theories predict an M-shaped recognition time function for the planar rotation of certain objects and provides a tractable mechanism for explaining the improved recognition performance typically recorded about 180° inversion (the restoration of categorical “side of” relations).

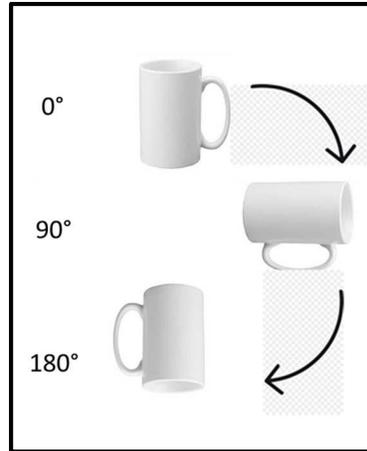


Figure 4. Coffee mug under planar rotation. The spatial relationship between the two parts of a coffee mug change from side-of to above-below at 90° planar rotation. However, at 180° the spatial relationship between these parts matches the side-of relationship shown at 0°.

In contrast, most template theoretical approaches predict a positive, monotonic relationship between recognition time and increasing planar rotation from a canonical viewpoint that would realize an upside-down V-shape. If energy-consuming mental rotation processes normalize the orientation of an imaged object to the stored view, called the “rotate-to-recognize” hypothesis, rotation functions should reach maximum and peak at 180° (Cooper & Shepard, 1973; Jolicoeur & Landau, 1984; Tarr, 1995; for subsequent experiments suggesting mental rotation is not used see Turnbull et al., 2002). Proponents of template theories, even while offering arguments opposed to structural descriptions, concede the dip presents a problem to simple templates (Jolicoeur, 1990). Researchers have occasionally removed the results obtained for rotations between 120° and 240° during analyses or excluded those manipulations in orientation from the outset, explaining the interval does not reliably index the mental rotation processes involved in recognizing objects at rotation (Jolicoeur, 1985; Large et al., 2003).

How mental rotation-based normalization processes could yield a bimodal function is unclear. Template theory proponents have speculated that a second parallel processing mechanism must exist to account for the dip. Interestingly, one author sympathetic to view-based

theory hypothesized that an additional recognition process encodes the relationship amongst an object's parts categorically, just as in structural descriptions (Jolicouer, 1985). The author speculated that the restoration of left/right categorical relations upon inversion causes the dip. This research has ironically been cited as strong evidence rendering structural description theory as an untenable approach (Gauthier & Tarr, 2016; Peissig & Tarr, 2006) despite its consistent predictions for the observed effects of rotation on recognition (Biederman, 1987; Hummel, 1994).

Perhaps supplying the most descriptive explanation for the dip using a template-based matching procedure, Murray (1997) explained that a dip in the rotation function for naming objects occurs because an exceptionally effective mental rotation strategy is afforded by inverted objects. Subjects instructed to mentally spin inverted objects produce slower reaction times than those instructed to mentally flip inverted objects (i.e. perform a top-bottom reflection). Therefore, Murray proposed that individuals spontaneously adopt the flipping strategy when objects appear at 180° while a mental rotation is performed at all other orientations, and an M-shaped rotation function precipitates as a result. The “flip-to-recognize” hypothesis does not have a reason to predict that planar rotation function shape should vary depending upon the structure of objects as was observed in Large et. al (2003). Large and colleagues used a naming procedure to examine the interaction between an object's axis of elongation, planar rotation, and recognition performance. The authors determined the direction of axis elongation and the presence of an elongated axis has no effect on the shape of planar rotation functions, except for that horizontally elongated objects (i.e. wide) were processed more efficiently at 180° than either vertically elongated (i.e. tall) and non-elongated objects.

CHAPTER 2. EXPERIMENT ONE

Hypotheses

The purpose of the first experiment is to determine whether the restoration of “side of” relationships accounts for the dip in response times at 180° observed for identifying objects rotated in the picture plane. If the dip at 180° is due to the restoration of “side of” relations, then objects with only “side-of” relations, like a coffee mug, should demonstrate the typical M-shaped reaction time function. In contrast, objects that only have “above” or “below” relations among their parts (like the suitcase in Figure 5) would not be expected to show the dip because they have no side of relations to restore.

Two different theoretical processes have been proposed by which template representations could be matched with the image of a rotated object. The “rotate-to-recognize” hypothesis predicts rotation functions for both “side-of” and “above/below” objects will demonstrate a peak at 180° because of the linear costs associated with mental rotation. The “rotate-to-recognize” hypothesis would be supported by upside-down V-shaped rotation functions observed for each object type. The “flip-to-recognize” hypothesis predicts rotation functions for both side-of objects and above/below objects will demonstrate a local minimum at 180° because planar inversion uniquely affords the spontaneous “flipping” of an object’s image and facilitates template matching. The “flip-to-recognize” hypothesis would be supported by obtaining M-shaped rotation functions for both object types (see Figure 6 on page 25).

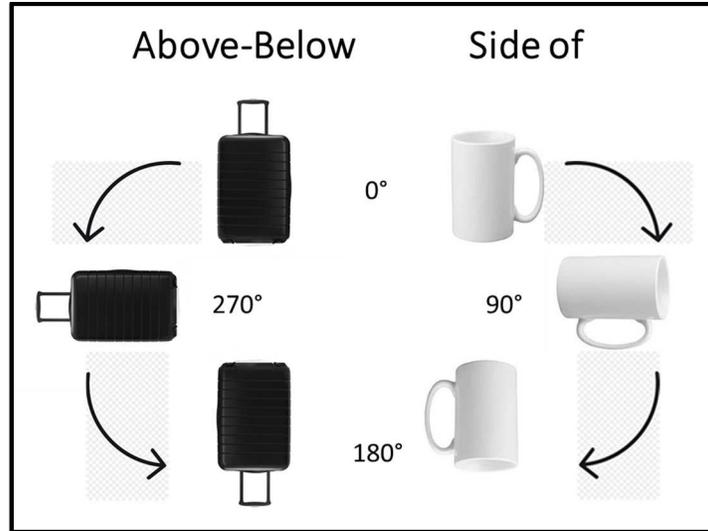


Figure 5. An Above-Below and a Side-of Object Rotated. This figure illustrates how rotation affects the spatial organization of the parts of a suitcase (an above-below object) and a coffee mug (a side-of object). For both objects, a 90° planar rotation changes the categorical spatial relations among parts. However, at 180° rotation these categorical relations are restored for the coffee mug while the above-below relations are reversed for the suitcase.

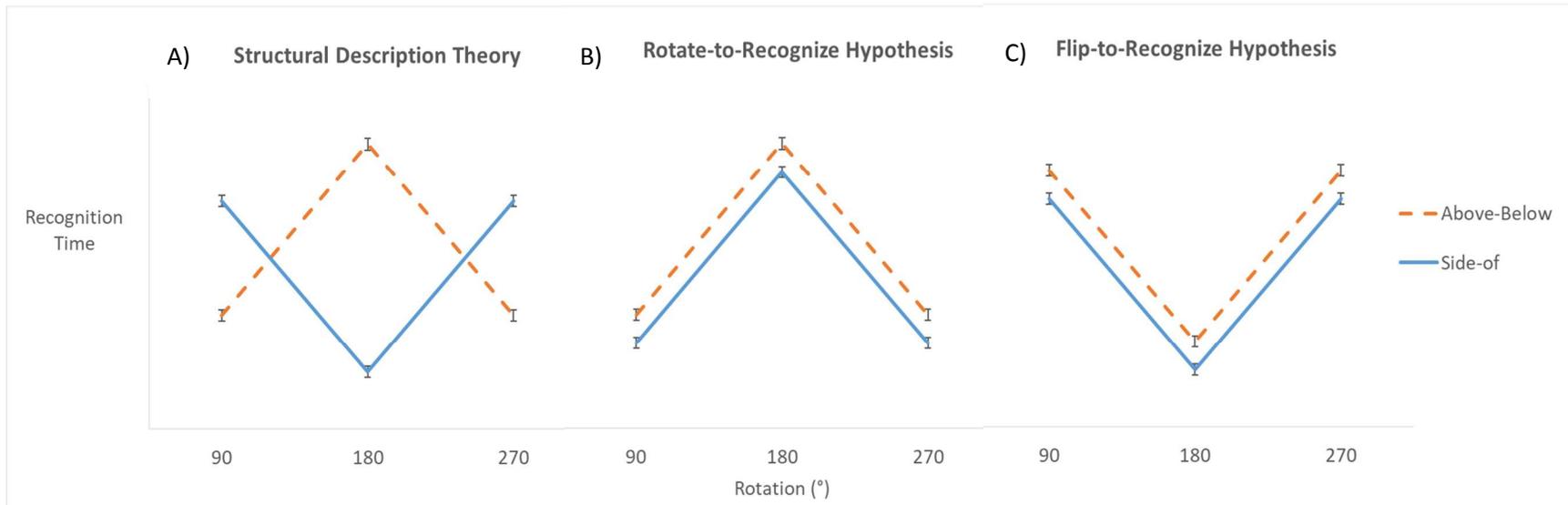


Figure 6. Predicted Rotation Function Shape for Above-Below and Side-of Objects by Theory. Theoretical predictions for reaction time functions are displayed, although these patterns also apply to the error rates anticipated for each theory. A) Structural Description Theory predicts that the restoration of side of categorical relations upon inversion will result in a local minimum at 180° for side-of objects only, while above-below should demonstrate an upside-down V-shaped function. B) The Rotate-to-Recognize Hypothesis predicts that recognition times increase monotonically for all objects regardless of their spatial relations. Therefore, an upside-down V-shaped function should occur for both side-of and above-below objects. C) The Flip-to-Recognize Hypothesis predicts that the restoration of an object's canonical axis upon inversion facilitates recognition. This effect theoretically occurs for any given object recognized at the basic-level, resulting in a V-shaped function for both object types.

Materials and Methods

Stimuli

Three types of objects were chosen for the experiment: objects with only side of relations, objects with only above-below relations, and non-diagnostic objects that had arbitrary types of relations (see Figure 7 for examples). Only objects with a canonical orientation were included because the identification of poly-oriented objects does not vary with rotation (Gibson & Robinson, 1935). The non-diagnostic category contained objects that were chosen arbitrarily as these stimuli are included only as distracters that prevent the participant from anticipating the orientation of the stimulus. Each object was presented in isolation, one at a time. Each image was edited using GIMP 2.0 for grayscale conversion and for regularization of scale, at approximately 300 pixels². Using grayscale images prevented participants from using color to make their identifications (Bramão, et. al, 2011). Local diagnostic cues, such as button markings on a camera, were edited out of the image by filling in those regions with other parts of the object and made to look uniform. This editing procedure was necessary for three object images.

Twenty-four side-of objects, twenty-four above-below objects, and sixteen distractor objects were selected. Diagnostic objects were rotated at 90°, 180°, and 270°, while distractor objects appeared at the upright (0°) orientation. Three randomized lists were created using a random number generator in which each object exemplar appears only once at one of the three orientations: 90°, 180°, and 270° for side-of and above-below sets and 0° for the non-diagnostic set (see below for examples). These three lists were combined such that every object occurred three times in the experiment to increase the number of stimuli each participant saw at each rotation by object group. There was a total of 192 trials in the experiment. The first counterbalance was order-reversed so that the list appeared in the opposite sequence, creating a

total of two counterbalances. Each participant was presented with twenty-four objects for each of the three orientations of interest for both the “side of” and “above-below” object types and all sixteen objects in the non-diagnostic set at the upright orientation three times respectively.

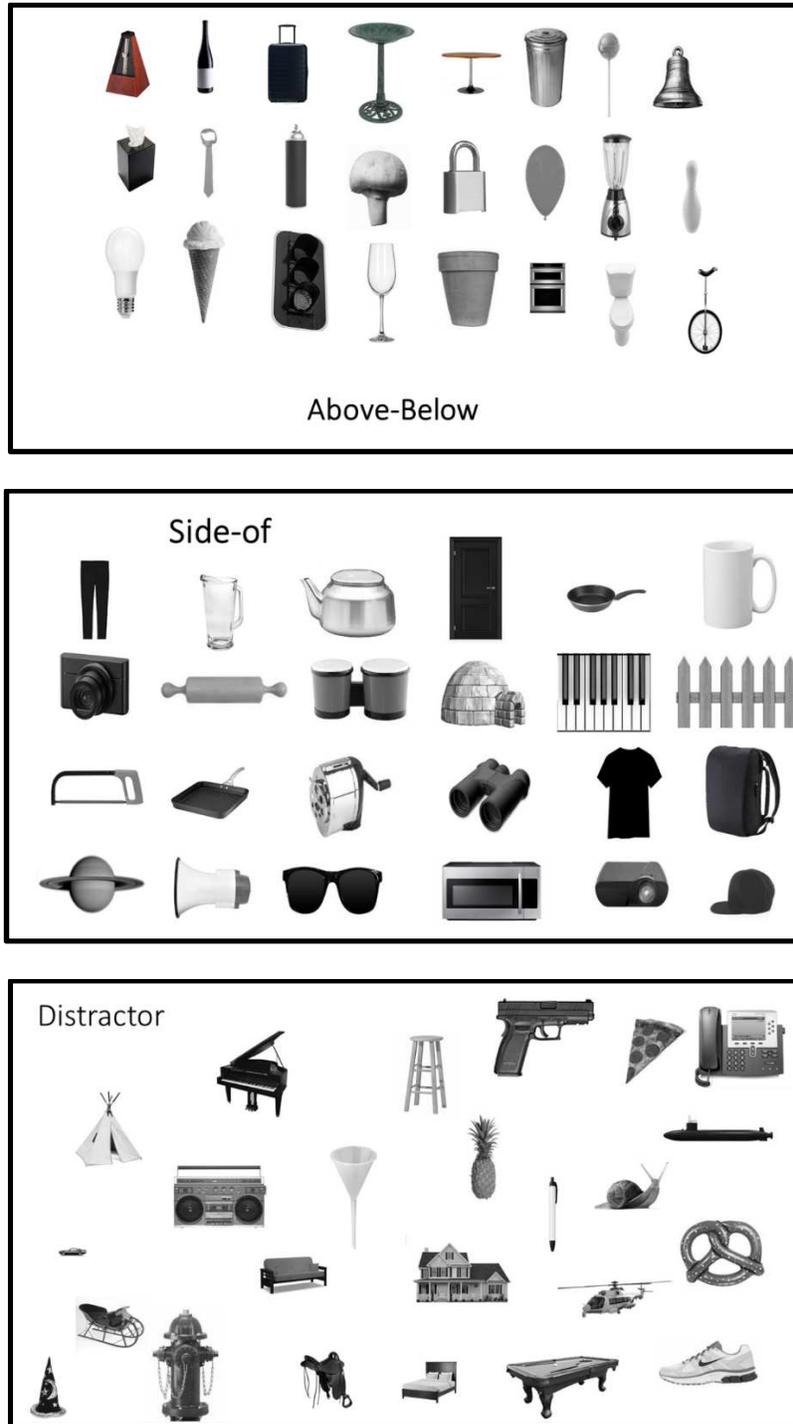


Figure 7. Experiment 1 Stimuli.

Methods.

A Dell desktop and monitor and E-Prime 2.0 software were used to conduct the experiment. Seventy-four participants were drawn from the Iowa State University subject pool and received course credit for their participation. Images were presented at half of a meter distance and were sized so that their maximum extent would fit into a $5^{\circ} \times 5^{\circ}$ of visual angle box. An experimenter and on-screen text instructed participants to say the name of the object when it appeared on the screen as quickly and accurately as possible on each trial. A fixation cross appeared for 150 milliseconds, followed by the object image for 300 milliseconds. After the object image disappeared, a visual mask was presented which was used to interfere with visual iconic memory for the preceding image of the object. The mask was rotated at random for each trial to prevent the participant from becoming desensitized to it. A microphone was used to detect vocal responses, and response times were recorded with the E-Prime software. The experimenter recorded whether the participant correctly named the object on each trial. Feedback was given to the participant after each trial regarding accuracy by displaying “Correct!” or “Incorrect.” and response time in milliseconds in order to provide motivation. A series of six practice trials with objects not present in the experimental portion was used to familiarize participants with the nature of the task before data collection.

Results

The results of Experiment 1 are shown in Figures 8 and 9. All statistical hypotheses reported in this paper were tested with a two-tailed alpha level of .05. False starts, distractor trials, and outliers were excluded from analyses. Outliers were defined as RTs outside 2.5 SDs from the mean for that stimulus (less than 5% of total trials). Within-participant factorial analysis-of-variance (ANOVAs) were conducted on the influence of object type (above-below

vs. side-of) and rotation (90°, 180°, 270°) on RT and error rates. Only accurate identifications were included in the analyses of RT.

RT data.

Neither the main effect of rotation, $F(2, 146) = 1.23, p = .30$, nor the main effect of stimulus type, $F(1, 73) = .11, p = .75$, were reliable. However, there was a reliable Stimulus Type X Rotation interaction, $F(2, 146) = 6.58, p = .002$. A planned reverse Helmert contrast compared the mean RT for side-of objects at 180° of rotation against the combined mean RT for side-of objects at 90° and 270°. This contrast was reliable, $F(1, 73) = 14.01, p = .001$. Thus, RT for side-of objects showed a local minimum at 180°. A second planned reverse Helmert contrast compared the mean RT for above-below objects at 180° of rotation against the combined mean RT for above-below objects at 90° and 270°. This contrast was not reliable $F(1, 73) = 2.00, p = .161$. There was no dip in RT for recognizing above-below objects.

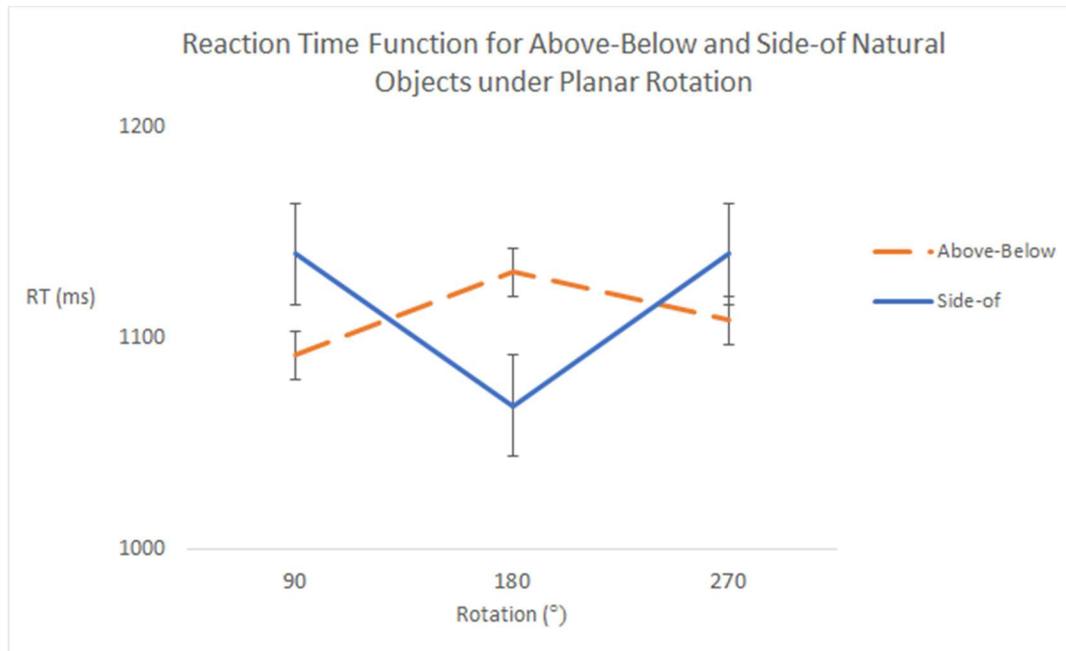


Figure 8. Experiment 1 planar function results for the time taken to name naturally occurring objects. Error bars represent standard error. A dip occurs for side-of objects and no dip occurs for above-below objects. A significant Object type x Rotation was observed.

Error data.

There was a reliable main effect of stimulus type, $F(1, 73) = 44.15, p < .001$. Participants made more errors recognizing above-below objects ($M = 11.9\%, SE = .012$) than they did recognizing side-of objects ($M = 7.9\%, SE = .009$). There was also a reliable main effect of rotation, $F(2, 146) = 6.33, p = .002$. Participants made fewer errors identifying objects at 180° ($M = 8.9\%, SE = .010$) than at 90° ($M = 10.5\%, SE = .010$) and 270° ($M = 10.4\%, SE = .010$). There was a Stimulus Type X Rotation interaction, $F(2, 146) = 6.66, p = .002$. A planned reverse Helmert contrast compared the mean error rate for side-of objects at 180° of rotation against the combined mean error rate for side-of objects at 90° and 270° . This contrast was reliable, $F(1, 73) = 34.39, p < .001$. Error rates for side-of objects showed a local minimum at 180° . A second planned reverse contrast compared the mean error rate for above-below objects at 180° of rotation against the mean error rate for above-below objects at 90° and 270° . This contrast was not reliable $F(1, 73) = 0.12, p = .73$. There was no dip in error rates for above-below objects.

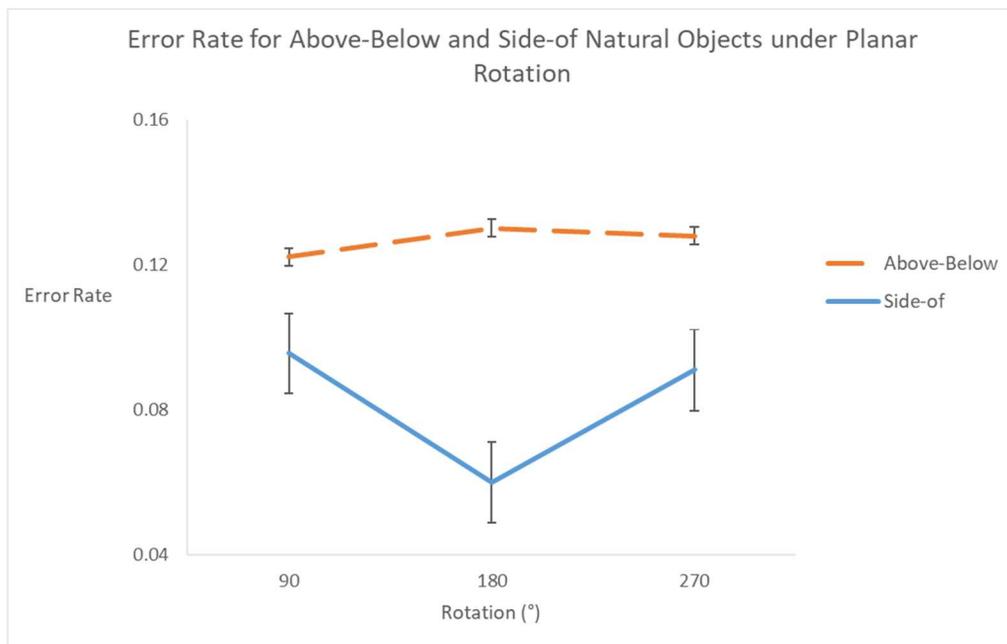


Figure 9. Experiment 1 planar function results for error rates naming naturally occurring objects. A dip occurred for Side-of objects while error rates showed no dip for Above-Below objects. Error bars represent standard error.

Discussion

The results of Experiment 1 suggest that rotation has a fundamentally different effect on the recognition of above-below objects and side-of objects. RTs and error rates for side-of objects demonstrated a dip at 180° of rotation, as predicted by categorical theories and the flip-to-rotate hypothesis. However, the “rotate-to-recognize” hypothesis would have no reason to expect improvements in performance at this orientation as it predicts the opposite pattern (a peak at 180°). In contrast, no dip was found in the RT and error rate rotation functions for above-below objects. The “rotate-to-recognize” and the “flip-to-recognize” hypotheses predict a peak or a dip should occur at 180°, respectively, which were not reflected in the observed rotation functions. A recognition theory positing categorical relations would not predict a dip to occur upon planar inversion of above-below objects because there are no side-of categorical relations to restore. Therefore, the results of Experiment 1 strongly suggest that the representations used to recognize basic-level objects encode part relations in categorical terms and that the M-shaped function typically observed in recognition studies using picture plane rotation results from the restoration of side-of categorical relations.

One alternative explanation for the results would assert the presence of confounding factors in the two object categories which influenced recognition outcomes nonrandomly. Of course, in experiments using naturally occurring objects, random assignment of stimuli to levels of the independent variable (i.e. groups) is impossible. One cannot randomly assign an everyday object to be “above-below” nor “side of”. However, if a similar experimental paradigm were to use constructed artificial objects that had been randomly assigned to their type of categorical relations, the issue of nonrandom assignment would be addressed. Such a paradigm was used in the second experiment.

CHAPTER 3. EXPERIMENT TWO

As Experiment 1 suggests, the belief that picture plane rotation findings challenge structural description theories as a tenable approach to object recognition may precipitate from a misunderstanding of their predictions. Perhaps subject to misinterpretation, the ambiguous phrase "object-centered" sometimes advanced renders them prone to criticism because of its implications for invariance. Object-centered recognition processes are considered fully independent from manners of object pose and are invariant in the ideal sense. However, Hummel and Biederman's JIM, a computer model with perceptual processes akin to those postulated in RBC theory, uses structural description formulae derived from an object's volumetric parts and the categorical relations amongst them, but importantly defines part relations such as above, below, and to the side in a manner *specific to the observer's viewpoint* (Hummel, 1994). JIM recognizes left-right reflections after a single exposure and demonstrates a local jump in performance when presented with objects at 180° (corresponding to both mirror generalization and the RT and error rate dips for human receivers).

The purpose of Experiment 2 was to replicate the results of Experiment 1 using artificial three-dimensional stimuli and a novel sequential same/different paradigm. Other studies have examined the effect of rotation on matching performance for sequentially presented artificial stimuli. However, most of these experiments either did not use stimuli which readily avail themselves to a structural description or used distractor objects which did not have a different structural description from the learned object. Arnoult (1954) used terrain-like two-dimensional amorphous blobs, Edelman and Bülhoff (1992) used a series of twisted tube and amoeba-like objects, and Dixon and Just (1978) used ellipses and color patches. These types of stimuli differ along a metric continuum rather than in structural descriptions and putatively engage template-

like perceptual processes. No dip upon inversion was observed in these studies as a result.

Experiment 2 addresses this issue by utilizing artificial stimuli created from combining three geons which can be discriminated by their structural description representation (see Figure 10).

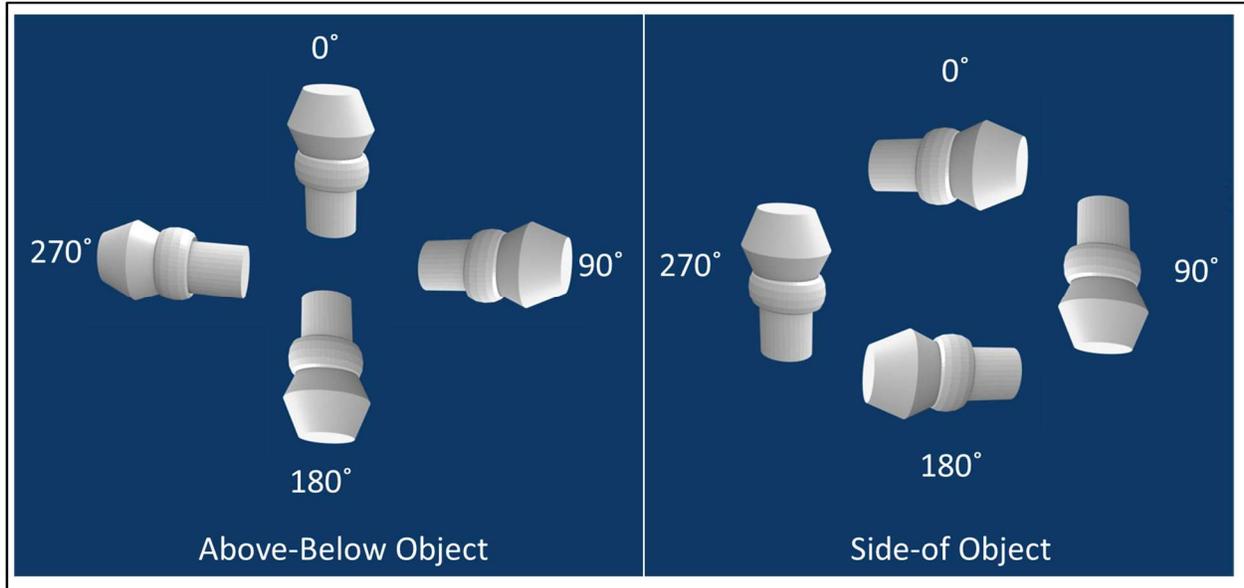


Figure 10. Above-below and side-of object with consistent geometry under rotation.

Artificial objects are pragmatic tools for studying misorientation's effect on recognition outcomes because participants have never encountered the stimuli a priori, and so the perspectives that participants use when formulating their visual representations are strictly controlled and can be manipulated. Rock (1973) observed that a diamond differs from a square only by 45° rotation. Similarly, the same set of geons connected in a consistent order may evoke structural representations with distinct properties depending on its orientation in the familiar view. Here, each stimulus was presented at an orientation ensuring all the categorical relationships between parts were either "side-of" or "above-below". Thereafter, the same stimuli made of the same parts arranged in a consistent order was presented again such that the alternate categorical part relations were afforded by its image (once above-below, once side-of, and vice

versa; see Figure 9), thereby negating the possibility for confounding differences between category membership by equating them, and enabling more scrupulous examination of spatial relationship effects, planar rotation effects, and their interaction.

Hypotheses

The purpose of Experiment 2 is to test the hypotheses posed in Experiment 1 using novel three-dimensional objects and a sequential same/different paradigm. Experiment 1 supported the notion that “side of” relationships account for the dip in response times and error rates at 180° rotation in the picture plane. If the dip is due to the restoration of “side of” relations, then objects presented with side-of relations during the learned view should demonstrate the typical M-shaped reaction time function. In contrast, objects that only have above-below relations among their parts would not be expected to show the dip because they have no “side of” relations to restore.

As described earlier, the “rotate-to-recognize” hypothesis predicts rotation functions for both “side-of” and “above/below” objects will demonstrate a peak at 180° because of the linear costs associated with mental rotation. The “rotate-to-recognize” hypothesis would be supported by upside-down V-shaped rotation functions observed for each object type. The “flip-to-recognize” hypothesis predicts rotation functions for both “side-of” objects and “above/below” objects will demonstrate a local minimum at 180° because planar inversion uniquely affords the flipping of an object’s image thus facilitating template matching. The “flip-to-recognize” hypothesis would therefore be supported by obtaining M-shaped rotation functions for both object types (refer to Figure 6).

Materials and Methods

Stimuli

The objects used in this experiment were created from a combination of three geons (see Figure 1 for individual geon depictions) using Blender 2.79 open-source image-rendering software. There were two groups of objects: priming objects and test objects.

Figures 11 and 12 on the next page provide examples of the stimuli used. Each of the novel three-dimensional objects displayed in the experiment were created by combining three different geons. Two geons were attached to a central geon at opposite ends relative to each other. This arrangement created totem pole-like objects with parts stacked only vertically or horizontally. Limiting the number of compositional geons to three was desirable because structural description theories such as JIM predict greater confusability for totem pole objects with four or more geons (Hummel, 1992). In a four geon totem pole, the two central geons share spatial descriptors (both above and below or side of other geons at either end). Therefore, distractor objects with more than three geons would have been overly difficult to discriminate. Using any less than three geons to compose the novel objects was also unfavorable because the matching task would be trivially easy. For distractor objects of the second variety, those with the same geons presented in a different spatial arrangement than that of the test object, the central geon always interchanged positions with either of the two end geons. This approach was necessary because exchanging the positions of two end geons results in an object that could also be created by rotating the original object 180° thereby conflating spatial relations and rotation. Aspect ratios, orientation of the long axes, and the surface region of connection were preserved when creating the distractors.

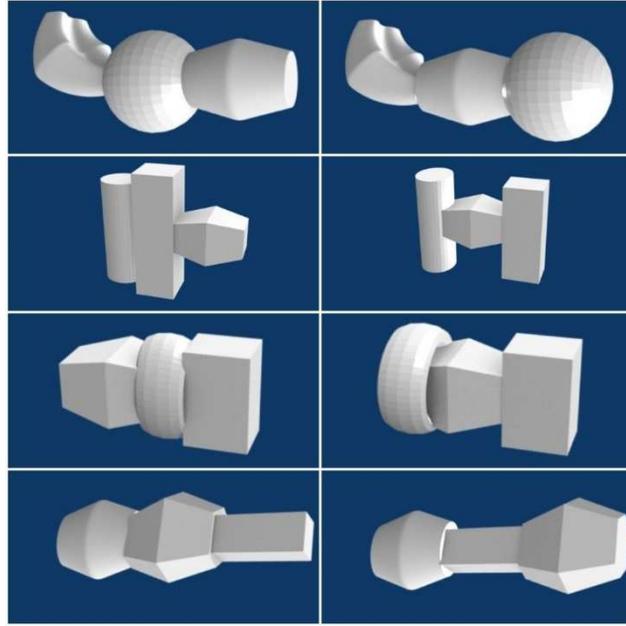


Figure 11. Examples of Different Object Sets with Two Geon Positions Exchanged. Each pair displayed has the same parts, although the two of the part positions have been altered for each object.

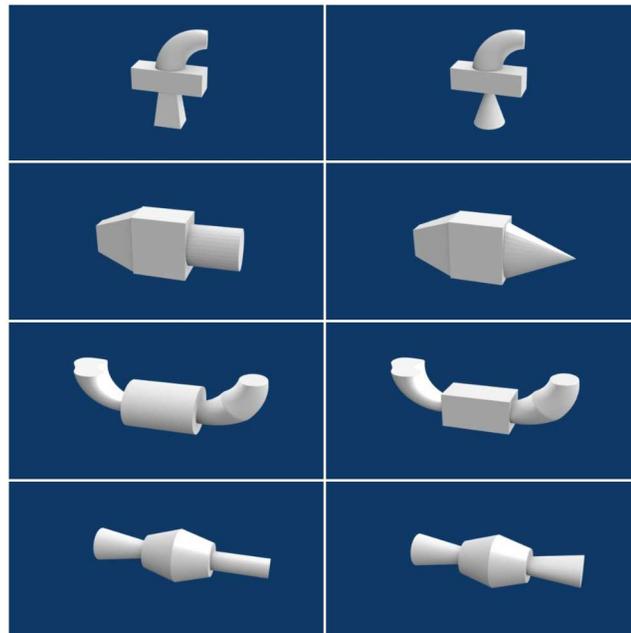


Figure 12. Examples of Different Object Sets with a Geon Swapped. For each of these object pairs, one geon is different between the two while the organization of parts is identical.

Methods

A Dell desktop computer and monitor and E-Prime 3.0 software were used to present object images. Fifty-one participants were drawn from the Iowa State University SONA subject pool and received course credit in return for their participation. On-screen instructions instructed participants of the nature of the task and to respond as quickly and accurately as possible. Participants were told that if the second object presented in a trial has the same parts and arrangement as the priming object for that trial, to press the “s” key for same even if the second presentation has been rotated. Likewise, if the second object presented in a trial is different from the first object, to press “n” for different. The sequence of events is illustrated in Figure 12.

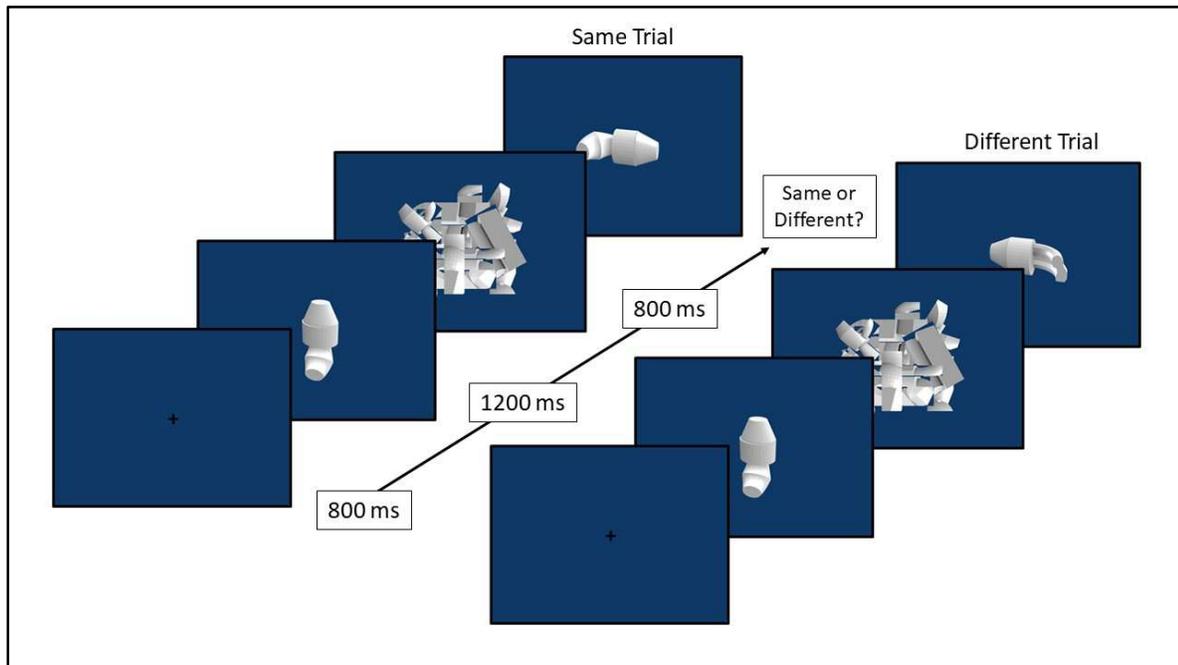


Figure 13. Experiment 2 procedure for same and different trials. The participant decided whether the second object matched the first object presented. A visual mask was displayed in between these object presentations.

Each object was presented in isolation, one at a time, following a fixation cue to focus visual attention for 800 milliseconds. The viewing distance was half of a meter, and the object stimuli were sized so that their maximum extent would fit in a $5^{\circ} \times 5^{\circ}$ of visual angle box. This priming object had either “above-below” relations or “side-of” relations and appeared for 1200 milliseconds. Then, participants were presented with a visual mask made of scrambled object parts used in the experiment for 800 milliseconds to disrupt iconic memory. At the test slide, either the same object was presented again (50% of trials) or a different object (50% of trials). As described, the different objects were identical to the primed object except that the position of two parts had been switched (always the middle geon with one of the two end geons) or one of the parts was different (see Figures 10 and 11 for examples of both). The second object for each trial appeared offset to the upper left, upper right, lower left, or lower right of the screen and were rotated at either 0° , 90° , 180° , or 270° on an equal number of trials. The test slide remained on the screen until the participant pressed either “s” for same object or “n” for different object. Response accuracy and time was recorded by E-Prime software. Participants were given feedback regarding their selection (“You indicated Same” or “You indicated Different”) and accuracy (“Correct!” or “Incorrect.”) to provide motivation.

There was a total of 272 trials in the experiment. Participants were given a break after 136 trials, and there were two counterbalanced lists used in the experiment that alternated order between participants. Each participant was presented with seventeen objects for each of the four orientations of interest during same trials for both the “side-of” and “above-below” object types.

Results

Distractor trials and outliers were excluded from analyses. Outliers were defined as RTs outside 2.5 SDs from the mean for that stimulus (less than 5% of total trials). Within-participant analysis-of-variance (ANOVAs) were conducted on the influence of object type (above-below vs. side-of) and rotation (0° , 90° , 180° , 270°) on RT and error rates. Only accurate same trials were included in the analyses of RT.

RT data.

The reaction time data is displayed in Figure 14 on page 40. There was a reliable main effect of rotation, $F(3, 150) = 39.80, p < .0001$. Overall, objects were recognized fastest at 0° rotation ($M = 1,023$ ms, $SE = 37.82$), followed by 90° ($M = 1,154$ ms, $SE = 37.46$), 270° ($M = 1,207$ ms, $SE = 40.71$), and 180° ($M = 1,247$ ms, $SE = 41.14$). No reliable main effect of stimulus type, $F(1, 150) = 1.57, p = .22$, was found. However, there was a reliable Stimulus Type X Rotation interaction, $F(3, 150) = 7.15, p < .001$. A planned Helmert contrast compared the mean RT for side-of objects at 180° of rotation against the combined mean RT for side-of objects at 90° and 270° . This contrast was not reliable, $F(1, 50) = 0.19, p = .669$. Thus, RT for side-of objects did not demonstrate a local minimum at 180° . Exploratory post-hoc trend analysis revealed a linear component, $F(1, 50) = 20.63, p < .001$, and a quadratic component, $F(1, 50) = 23.19, p < .001$, from 0° to 180° (with 90° and 270° collapsed) for the side-of function. A second planned Helmert contrast compared the mean RT for above-below objects at 180° of rotation against the combined mean RT for above-below objects at 90° and 270° . This contrast was reliable $F(1, 50) = 22.30, p < .001$. There was a peak in RT for recognizing above-below objects observed at 180° . Another exploratory post-hoc trend analysis revealed a linear component, $F(1,$

50) = 95.17, $p < .001$, and no quadratic component for the above-below object rotation function, $F(1, 50) = 0.003$, $p = .96$, from 0° to 180° (with 90° and 270° collapsed).

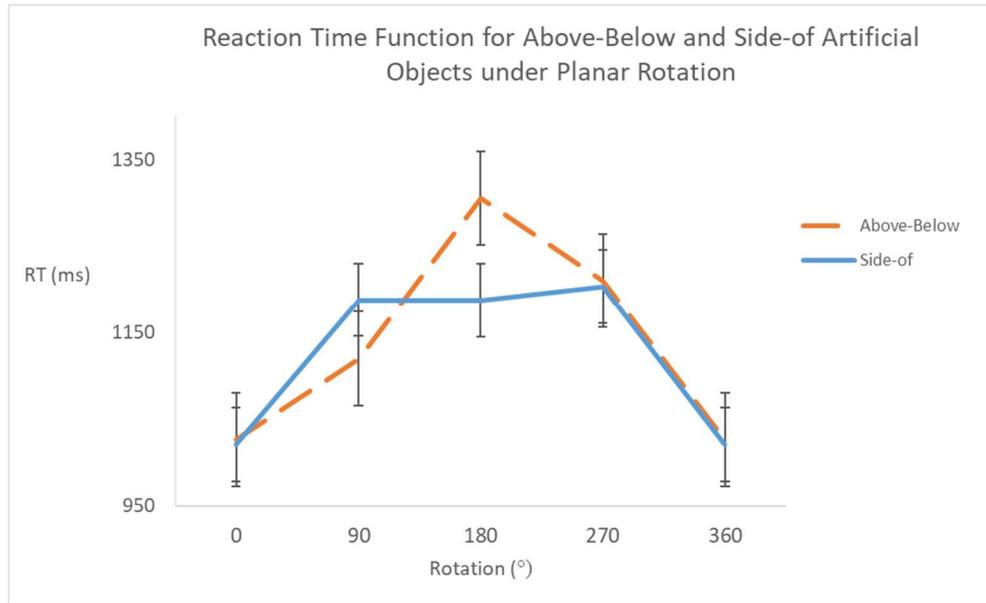


Figure 14. Reaction time function for above-below and side-of artificial objects under planar rotation. Data points at 0° repeated at 360° for symmetry. Error bars represent standard error.

Error data.

The error rate data is displayed in Figure 15 on page 41. There was a reliable main effect of stimulus type, $F(1, 50) = 10.311$, $p = .002$. Participants made fewer errors recognizing above-below objects ($M = 20.6\%$, $SE = .013$) than they did recognizing side-of objects ($M = 23.6\%$, $SE = .015$). There was also a reliable main effect of rotation, $F(3, 150) = 69.29$, $p < .0001$.

Participants made fewer errors identifying objects at 0° ($M = 9.6\%$, $SE = .014$) than at 90° ($M = 24.1\%$, $SE = .015$), 180° ($M = 24.4\%$, $SE = .020$), and 270° ($M = 30.2\%$, $SE = .016$). There was a Stimulus Type X Rotation interaction, $F(3, 150) = 3.62$, $p = .015$. A planned Helmert contrast compared the mean error rate for side-of objects at 180° of rotation against the combined mean error rate for side-of objects at 90° and 270° . This contrast was reliable, $F(1, 50) = 10.01$, $p < .01$.

Thus, error rates for side-of objects showed a local minimum at 180°. A second Helmert planned contrast compared the mean error rate for above-below objects at 180° of rotation against the aggregated mean error rate for above-below objects at 90° and 270°. This contrast was not reliable $F(1, 50) = .001, p = .981$. There was no dip in error rates for recognizing above-below objects.

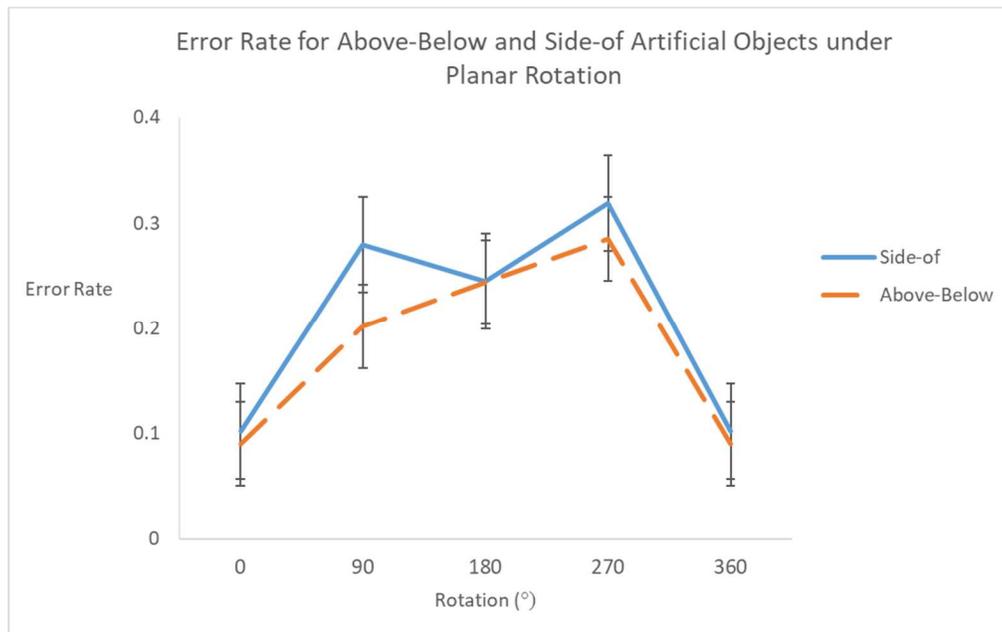


Figure 15. Error function for above-below and side-of artificial objects under planar rotation. Data points at 0° repeated at 360° for symmetry. Error bars represent standard error.

Discussion

In Experiment 2, rotation affected recognition performance differently when objects were composed of parts with only above-below relations than when they had only side-of relations. The reaction time rotation function peaked for above-below objects and leveled off for side-of objects at 180°. The error rate rotation function leveled off for above-below objects and “dipped” for side-of objects at 180°. This robust interaction is particularly counterintuitive because the

object categories contained the exact same stimuli with the only difference between the two groups being the manner in which the objects were oriented during the priming trial.

Interestingly, the reaction times between each group were only significantly different from one another when the objects were inverted in the picture plane. Side-of objects were identified faster than above-below objects at the 180° locus. None of the recognition approaches predicted that the reaction time rotation function for above-below objects should peak while the rotation function for side-of objects should level off at inversion. However, structural description theories would have the easiest time describing this outcome because they predict inversion should affect the recognition of side-of objects less than above-below objects because inversion restores side-of relationships. No other theoretical approach would predict an interaction between the stimulus type and the effect of rotation. Moreover, the error rate function did demonstrate a dip 180° for side-of objects and did not for above-below, providing support for the notion that the restoration of categorical relations upon inversion explains the dip in rotation functions.

A potential explanation for the rotation function leveling off for side of objects instead of dipping could be that different recognition processes were used than those used in naming experiments. No previous experiment has found a dip in reaction times for basic-level object matching tasks using planar rotation as a manipulation. Biederman and Gerhardstein (1993) suggest that old/new recognition tasks do not permit access to underlying view-independent processes as well as naming tasks. Therefore, a dip may not be the expected outcome for this experiment per se, although side-of objects should be relatively easier to recognize when inverted than above-below objects.

CHAPTER 4: GENERAL DISCUSSION

The purpose of the preceding experiments was to determine the cause of the counterintuitive improvement in recognition when objects are rotated 180°. The first study used everyday objects and demonstrated a clear difference in the rotation function occurring between 90° and 270° for objects with a vertical and horizontal arrangement of parts. The second study used artificial objects to control for potential confounds present in Experiment 1 and replicated the interaction. The effect of planar rotation on recognition depends on the spatial relations among an object's parts. Progressively poorer recognition for all object groups with increasing angular disparity would have supported the "rotate-to-recognize" hypothesis and observing a dip for all object groups would have supported the "flip-to-recognize" hypothesis. However, rotation functions did not adhere to either of these two outcomes. Instead, recognition performance seemed acutely sensitive to the type of categorical relations an object portrayed when its image was rotated. Recognition was facilitated when categorical relations were consistent with an object's stored structural description and hindered when they were inconsistent. When side-of relations were restored upon inversion, this match reduced the negative impact of rotation on recognition. Conversely, the presence of mismatching categorical relations (particularly for above-below objects) appeared to be deleterious for recognition as the typical dip was replaced by a peak in some of the measures employed here.

A limitation of the second experiment comes from the assumption that the orientation of a once-presented object becomes the accepted canonical view. Rather than assume passive acceptance of the offered orientation as the preferred comparative, canonical orientations may be influenced by those positions at which the object is in gravitational equilibrium (Leone, 1998). Indeed, select examples appeared more physically stable at certain rotations than others. Future

research should delineate whether properties of an object's three-dimensional shape that imply a canonical orientation other than the orientation of the object in the provided image influence the nature of the representation developed for that object.

With the establishment of corresponding rotation functions between objects of above-below and side-of spatial relations respectively in both experiments, future experiments can examine aspects of object shape using the sequential matching paradigm which were not considered here. For example, open questions exist regarding whether non-connected part relations are encoded in structural descriptions and how parts with both above-below and side-of relations are represented. Many everyday objects exhibit a combination of vertical and horizontal spatial relations between their compositional parts. How the presence of both categorical positionings within an object influences recognition under picture plane rotation in concert remains an open question. Though side-of relations are restored for these objects upon inversion, above-below spatial relations would simultaneously appear opposite the stored description. Whether the presence of discordant spatial information overrides, matches, or is overridden by the presence of concordant spatial descriptors could be investigated with this paradigm. Many other manipulations are imaginable, such as examining how the number of an object's geons impacts response times and accuracy rates for detecting mismatches, the influence of connectedness to a central part, and testing the effect of rotation in depth on recognition measures. A future experiment could also test whether repetition blindness occurs for side-of objects that have been rotated 180° more often than repetition blindness for above-below objects.

Structural Descriptions appear to be especially well equipped to explain the unexpected patterning of naming times for basic-level object identification of rotated objects. An important note to include here is that finding evidence for the operation of a recognition system under

certain experimental conditions does not definitively negate the potential existence of a secondary processing system which operates in parallel, simultaneously. An alternative explanation could assert duality, whereby each processing technique handles certain tasks more efficiently than the other and could even function in cooperation. In fact, many modern object recognition theories include both structural and template-like pathways (Burgund & Marsolek, 2000; Cooper & Wojan, 2000, Hummel, 2013). Depending on the task specific demands, evidence for either processing operation often surfaces (Gauthier & Tarr, 2016).

The results from this experiment explain peculiar phenomenon obtained in many previous studies. Harris and Dux (2005) employed a repetition blindness paradigm during which sequentially presented images of the same object were only noticed as separate instantiations if the second critical item was flipped upside down. When the second image was rotated in the picture plane at any rotation other than 180° , participants were much less likely to report seeing a second image of the object. Based on these results, it appears that inversion produces an image most different from that of the upright. Interestingly, most of the objects presented in this experiment had a vertical axis of elongation, which would be reestablished upon inversion. Thus, above/below relations become reversed, and RBC would predict that the inverted object activates a maximally contrasting representation, in terms of categorical spatial relations, enabling rapid detection of the novel presentation. Davidoff and Warrington (1999) report a patient whom could differentiate between upright and 180° rotated objects that had a base but could not differentiate between the same objects at rotations other than 180° . This suggests that orientation blindness can be overcome when objects with vertical alignment assume a maximally different structural description, or in other words, when above-below relations become reversed in an object.

There are practical applications that follow from the present research regarding human visual disorders. Attaining expertise in reading and writing appears greatly contingent upon the suppression of mirror generalization processes in literary contexts (Ahr, Houdé, & Borst, 2016). Furthermore, the skill of inhibiting mirror generalization in reading stimulates the general ability to discriminate between lateral mirror images of many other visual stimuli. Preliterate children, illiterate adults, and monoliterate individuals whose language is devoid of orthographic mirror images have more difficulty discriminating between left-right reflections of an image and between planar inversions of an object than literate adults (Kolinsky et al., 2011). Planar inversions result in an image that has been both left-right mirror reflected and top-bottom reflected.

Alarming, some researchers argue that most reading disorders result from the unusual perceptual demands placed on our recognition system by the orthography in written languages (Taeko & Brian, 1999). Indeed, the present research argues that coding horizontal relations categorically as “side-of” without distinguishing left and right (mirror generalization) is an integral function of our basic object recognition processes. A mistake commonly associated with dyslexia involves confusing the sequence of the letters within a word (e.g. *was* and *saw*, *pat* and *tap*, etc.) which would be anticipated during the operation of a recognition system without explicit provisions for specifying handedness. Cultures with written languages containing enantiomorphs or mirror-reversed letters such as “p” and “q” and “d” and “b” report ten times higher rates of childhood dyslexia than those languages without enantiomorphs (Makita, 1968). A vast amount of research has been conducted aimed to develop treatments or font-styles that are easier for dyslexics to read. Despite the promise offered by these products, empirical evidence suggests that both dyslexics and non-dyslexics are unfortunately no better at reading with these

fonts than they are at reading Times New Roman or Arial (Kuster et al., 2017). The research described here may offer critical insight into the proper design of texts that facilitates reading for dyslexics.

Using what we know about structural description representations, the root cause of the trouble dyslexics face while reading could be that words are processed using basic-level object recognition centers instead of a distinct area designated for word recognition. Because the object recognition centers do not readily differentiate relative left and right positions, letters may be perceived in the wrong order. However, a text which is presented vertically may be less susceptible to inversions because “above” and “below” are coded as discrete non-overlapping categories. Indeed, Lee et al. (2002) found that when text was arranged vertically, a patient who sustained brain injury leading to dyslexia committed less mistakes than when she was reading text arranged horizontally. Additionally, rates of childhood dyslexia are diminished in languages like Japanese that write from top to bottom (i.e., the direction in which RBC theory predicts inversions are very unlikely to occur) and that write in both directions, using mirror reflections (Makita, 1968)

A growing number of computer vision approaches are incorporating explicit code for identifying visual features and their spatial relationship when identifying objects in an image. Deformable part-based models such as these have proven superior to bag-of-feature models which do not consider the spatial relationship amongst an object’s defining features. The most successful attempts to solve this problem utilize processing steps comparable to those used in matching structural descriptions. For example, part-based models represent objects as concatenates of parts and flexible spatial relations (Fischler & Elschlager, 1973 provided the original framework). Recent versions can successfully classify objects in test images after an

entirely unsupervised learning phase. That is, no feedback is given to the model about whether it has produced a correct response, yet it will still efficiently encode the consistencies in an object's features and their relative size and locations. Moreover, part-based models require a minimal number of training images, handle partial occlusion, function in cluttered environments, and are invariant to affine transformations (Zhang, 2014). Thus, the research presented here contributes to our understanding of human object recognition and could fuel the development of more robust computer vision algorithms.

References

- Ahr, E., Houdé, O., & Borst, G. (2016). Inhibition of the mirror generalization process in reading in school-aged children. *Journal of Experimental Child Psychology*, *145*, 157–165. <https://doi.org/10.1016/j.jecp.2015.12.009>
- Amir, O., Biederman, I., & Hayworth, K. J. (2012). Sensitivity to nonaccidental properties across various shape dimensions. *Vision Research*, *62*, 35–43. <https://doi.org/10.1016/j.visres.2012.03.020>
- Arguin, M., & Saumier, D. (2004). Independent processing of parts and of their spatial organization in complex visual objects. *Psychological Science*, *15*(9), 629–633.
- Arnold, L. F. (2014). *Does categorical perception interfere with drawing accuracy?*, Graduate Theses and Dissertations.
- Arnoult, M. D. (1954). Shape discrimination as a function of the angular orientation of the stimuli. *Journal of Experimental Psychology*, *47*(5), 323–328. <https://doi.org/10.1037/h0060065>
- Biederman, I. (1987). Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review*, *94*(2), 115-147.
- Bramão, I., Reis, A., Petersson, K. M., & Faisca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *ACTPSY*, *138*, 244–253. <https://doi.org/10.1016/j.actpsy.2011.06.010>
- Burgund, E. D., & Marsolek, C. J. (2000). Viewpoint-invariant and viewpoint-dependent object recognition in dissociable neural subsystems. *Psychonomic Bulletin & Review*, *7*(3), 480–489. <https://doi.org/10.3758/BF03214360>
- Rosielle, L., & Cooper, E. (2001) Categorical perception of relative orientation in visual object recognition. *Memory & Cognition*, *29*(1), 68-82.
- D’Innocente, A., Carlucci, F. M., Colosi, M., & Caputo, B. (2017). Bridging between computer and robot vision through data augmentation: A case study on object recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10528 LNCS*(figure 2), 384–393.
- Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel view of three-dimensional objects. *Vision Research*, *32*(12), 2385–2400.
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). *CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms*. (Iccc), 1–22. Retrieved from <http://arxiv.org/abs/1706.07068>
- Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Research*, *41*(2), 221–234.
- Gauthier, I, Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience* *2*, 6. *June*, *2*(6), 568–573.

- Gauthier, Isabel, Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2002). BOLD Activity during mental rotation and viewpoint-dependent object recognition. *Neuron*, *34*(1), 161–171.
- Gauthier, Isabel, & Tarr, M. J. (2016). Visual Object Recognition: Do We (Finally) Know More Now Than We Did? *Annual Review of Vision Science*, *2*(1), 377–396.
- Ghodrati, M., Farzmahdi, A., Rajaei, K., Ebrahimpour, R., & Khaligh-Razavi, S.-M. (2014). Feedforward object-vision models only tolerate small image variations compared to human. *Frontiers in Computational Neuroscience*, *8*(July), 1–17.
- Gibson, J. J., & Robinson, D. (2011). Orientation in visual perception; The recognition of familiar plane forms in differing orientations. *Psychological Monographs*, *46*(6), 39–47.
- González-Casillas, A., Parra, L., Martin, L., Avila-Contreras, C., Ramirez-Pedraza, R., Vargas, N., Ramos, F. (2018). Towards a model of visual recognition based on neurosciences. *Procedia Computer Science*, *145*(May), 214–231.
- Halwvlon, K. M. (1988). Cognition, 29 (1988) 29-46. *Science*, *29*, 29–46.
- Harris, I. M., & Dux, P. E. (2005). Turning objects on their heads: The influence of the stored axis on object individuation. *Perception and Psychophysics*, *67*(6), 1010–1015. <https://doi.org/10.3758/BF03193627>
- He, X., Yang, Z., & Tsien, J. Z. (2011). A Hierarchical Probabilistic Model for Rapid Object Categorization in Natural Scenes. *6*(5). <https://doi.org/10.1371/journal.pone.0020002>
- Hildreth, G. (1934). Reversals in Reading and Writing. *The Journal of Educational Psychology*, *XXV*(1). Retrieved from <http://psycnet.apa.org.proxy.lib.iastate.edu/fulltext/1934-02679-001.pdf>
- Hummel, J. E. (1993). *Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition*. <https://doi.org/10.1080/13506280143000214>
- Hummel, J. E. (2013). *Object Recognition*. 32–46. Retrieved from <https://pdfs.semanticscholar.org/0bce/853dcf4a35ba28a79b9e4374ad0a65e29b76.pdf>
- Jolicoeur, P., & Landau, M. J. (1984). Effects of orientation on the identification of simple visual patterns. *Canadian Journal of Psychology*, *38*(1), 80–93. <https://doi.org/10.1037/h0080782>
- Jolicoeur, Pierre. (1985). The time to name disoriented objects. ANGLE + PRACTICE. *Memory and Cognition*, *13*(4), 289–303.
- Jolicoeur, Pierre. (1990). *A Dual-systems Theory*. *5*(4).
- Kolinsky, R., Verhaeghe, A., Fernandes, T., Mengarda, E. J., Grimm-Cabral, L., & Morais, J. (2011). Enantiomorphy Through the Looking Glass: Literacy Effects on Mirror-Image Discrimination. *Journal of Experimental Psychology: General*, *140*(2), 210–238. <https://doi.org/10.1037/a0022168>
- Koriat, A., & Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 421–434. <https://doi.org/10.1037/0278-7393.10.3.421>

- Koriat, A., & Norman, J. (1985). Mental rotation and visual familiarity. *Perception & Psychophysics*, 37(5), 429–439. <https://doi.org/10.3758/BF03202874>
- Kr, N. (2003). *Biologically Motivated Multi-modal Processing of Visual Primitives*. 1(4).
- Kranjec, A., Lupyan, G., Chatterjee, A., & Bremner, A. (2014). Categorical Biases in Perceiving Spatial Relations. *PLoS ONE*, 9(5). <https://doi.org/10.1371/journal.pone.0098604>
- Kubilius, J., Schrimpf, M., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2018). CORnet: Modeling the Neural Mechanisms of Core Object Recognition. *BioRxiv*, 408385. <https://doi.org/10.1101/408385>
- Kuster, S. M., Van Weerdenburg, M., Gompel, M., & Bosman, A. M. T. (n.d.). *Dyslexie font does not benefit reading in children with or without dyslexia*. <https://doi.org/10.1007/s11881-017-0154-6>
- Large, M. E., McMullen, P. A., & Hamm, J. P. (2003). The role of axes of elongation and symmetry in rotated object naming. *Perception and Psychophysics*, 65(1), 1–19. <https://doi.org/10.3758/BF03194779>
- Leone, G. (1998). The effect of gravity on human recognition of disoriented objects. *Brain Research Reviews*, 28(1–2), 203–214. [https://doi.org/10.1016/S0165-0173\(98\)00040-X](https://doi.org/10.1016/S0165-0173(98)00040-X)
- Lowe, D. G., & Binford, T. O. (1983). Perceptual organization as a basis for visual recognition. *Proc. AAAI'83*, 255–260.
- Makita, K. (1968). the Rarity of Reading Disability in Japanese Children. *American Journal of Orthopsychiatry*, 38(4), 599–614. <https://doi.org/10.1111/j.1939-0025.1968.tb02428.x>
- Morton, N., & Morris, R. G. (1995). Image Transformation Dissociated from Visuospatial Working Memory. *Cognitive Neuropsychology*, 12(7), 767–791. <https://doi.org/10.1080/02643299508251401>
- Murray, J. E. (1997). Flipping and spinning: Spatial transformation procedures in the identification of rotated natural objects. *Memory and Cognition*, 25(1), 96–105. <https://doi.org/10.3758/BF03197287>
- O'Brien, A. M. (2018). Using Prosopagnosia to Test and Modify Visual Recognition Theory. *Perceptual and Motor Skills*, 125(1), 57–80. <https://doi.org/10.1177/0031512517745412>
- Peissig, J. J., & Tarr, M. J. (2006). *Visual Object Recognition: Do We Know More Now Than We Did 20 Years Ago?* <https://doi.org/10.1146/annurev.psych.58.102904.190114>
- Peissig, J. J., & Tarr, M. J. (2007). *Visual Object Recognition : Do We Know More Now Than We Did 20 Years Ago ?* <https://doi.org/10.1146/annurev.psych.58.102904.190114>
- Rice, C. (2019). *Linked references are available on JSTOR for this article : The Orientation of Plane Figures as a Factor in Their Perception by Children*. 1(2), 111–143.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3 Suppl, 1199–1204.

- Rolls, E. T., & Tovee, M. J. (2017). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, 73(2), 713–726.
- Rolls, Edmund T., & Mills, W. P. C. (2018). Non-accidental properties, metric invariance, and encoding by neurons in a model of ventral stream visual object recognition, VisNet. *Neurobiology of Learning and Memory*, 152(February), 20–31.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703. <https://doi.org/10.1126/science.171.3972.701>
- Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). *Increased knowledge about the ways people recognize each other may help to guide efforts to develop practical automatic face-recognition systems*. 94(11). <https://doi.org/10.1109/JPROC.2006.884093>
- Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: evidence for a dual representation of object shape. *Journal of Experimental Psychology. Human Perception and Performance*, 24(3), 732–744.
- Taeko, N. W., & Brian, B. (1999). A case study of an English-Japanese bilingual with monolingual dyslexia. *Cognition*, 70, 273–205.
- Tanaka, J. W., & Taylor, M. (1991). Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder? *Cognitive Psychology*, 23, 457–482. c
- Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of mental transformations in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, 2(1), 55–82.
- Tarr, M. J., & Kriegman, D. J. (2001). What defines a view? *Vision Research*, 41(15), 1981–2004. [https://doi.org/10.1016/S0042-6989\(01\)00024-4](https://doi.org/10.1016/S0042-6989(01)00024-4)
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2), 233–282.
- Thoma, V., Davidoff, J., & Hummel, J. E. (1987). Priming of plane-rotated objects depends on attention and view familiarity. *Hummel & Biederman*. <https://doi.org/10.1080/13506280500155627>
- Turnbull, O. H., Della Sala, S., & Beschin, N. (2002). Agnosia for object orientation: Naming and mental rotation evidence. *Neurocase*, 8(4), 296–305. <https://doi.org/10.1093/neucas/8.4.296>
- Turnbull, O. H., & McCarthy, R. A. (1996). When is a view unusual? A single case study of orientation-dependent visual agnosia. *Brain Research Bulletin*, 40(5–6), 497–502. [https://doi.org/10.1016/0361-9230\(96\)00148-7](https://doi.org/10.1016/0361-9230(96)00148-7)
- Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32(3), 193–254. [https://doi.org/10.1016/0010-0277\(89\)90036-X](https://doi.org/10.1016/0010-0277(89)90036-X)
- Wakui, E., Jü Ttner, M., Petters, D., Kaur, S., Hummel, J. E., Davidoff, J., & Watanabe, K. (2013). Earlier Development of Analytical than Holistic Object Recognition in Adolescence. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0061041>

- Wasserman, E. A., Kirkpatrick-Steger, K., Hamme, L. J. Van, Biederman, I., & Biederman, I. (1993). Pigeons Are Sensitive to the Spatial Organization of Complex Visual Stimuli. *Source: Psychological Science, 4186138*(5), 336–341. Retrieved from <http://www.jstor.org/stable/40063058>
- Wilson, K. D., & Farah, M. J. (2006). Distinct patterns of viewpoint-dependent BOLD activity during common-object recognition and mental rotation. *Perception, 35*(10), 1351–1366. <https://doi.org/10.1068/p5571>
- Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience, 11*(11), 1352–1360. <https://doi.org/10.1038/nn.2202>
- Yamins, D. L. K., 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 of the United States of America, 111*(23), 8619–8624. <https://doi.org/10.1073/pnas.1403112111>
- Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences, 10*(7), 301–308. <https://doi.org/10.1016/j.tics.2006.05.002>
- Zhang, W., & Science, C. (2014). *Constellation Models for Recognition of Generic Objects*. 1–10.