Stony Brook University

Exploring the Effect of Hierarchy on Categorical Search

A Dissertation Presented

by

Justin Taylor Maxfield

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Cognitive Science

Stony Brook University

May 2017

Stony Brook University

The Graduate School

Justin Taylor Maxfield

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Dr. Gregory J. Zelinsky
Departments of Psychology and Computer Science

Dr. Christian C. Luhmann
Department of Psychology

Dr. Nicholas R. Eaton
Department of Psychology

Dr. Corey J. Bohil
Department of Psychology
University of Central Florida

This dissertation is accepted by the Graduate School

Charles Taber

Dean of the Graduate School

Abstract of the Dissertation

Exploring the Effect of Hierarchy on Categorical Search

by

Justin Taylor Maxfield

Doctor of Philosophy

in

Cognitive Science

Stony Brook University

2017

This dissertation outlines the importance of understanding how hierarchically organized categories of objects are represented and used to complete novel tasks. Four experiments are included with the goal of expanding on the current field of research. These experiments explored the behavioral factors impacting visual search efficiency, the neural correlates of categorical search, and how to best model this behavior in order to make further testable predictions. Experiment one builds off of previous research done on how hierarchical categorical cues impact search by manipulating the set size of the search display. Experiment two focuses on identifying the characteristics of the N2pc EEG component when presented with hierarchical cues. Experiment three compares and contrasts the ability of the Category-Consistent Feature (CCF) model and Multi-dimensional scaling methods to predict performance on a category verification task. Experiment four extends the CCF model from predicting overall trends in behavioral data to making predictions about search efficiency at the level of individual trials. Throughout this dissertation, the importance of understanding more about these areas of research are highlighted. Given the ubiquity of categorical search in everyday life, there is a need to further our modelling efforts to generate new predictions and avenues of research.

Dedication Page

For Linnea.

Table of Contents

List of Figures/Tables/Illustrations

Acknowledgments

I would like to acknowledge the years of faithful work of my advisor Greg Zelinsky, without whom this work and the advancement of knowledge that it represents would not have been possible.

General Introduction

It is difficult to understate the importance of categories. The ability for humans to recognize objects, use previous knowledge to classify them and apply that information to novel situations is essential. From early human history, in discriminating which plants were safe to eat and which were not, we have been using our knowledge of categories on a daily basis. Categories form the very backbone of our psychological being. They grow with us through our lives and give structure to the concepts that define who we are (Kaplan & Murphy, 2000; Murphy, 2002; Pazzani, 1991). Therefore, it is unsurprising that a great deal of psychological research has been dedicated to understanding the cognitive principles that underlie categories, from learning them (Ashby & Maddox, 1993; Kruschke, 1992; Love, Medin, & Gureckis, 2004; Nosofsky, 1986), remembering them (Anderson, 1983, Nosofsky, Cao, Cox, & Shiffrin, 2014), and using them every day.

One way of organizing categories of objects is to group them hierarchically. Any given object can be categorized at multiple levels in such a hierarchy (Mack & Palmeri, 2011; Murphy, 2002; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). For example, a sea vessel powered by the wind can be called a sail boat (subordinate), simply a boat (basic), or more broadly a vehicle (superordinate). There is a loose set of rules regarding this hierarchical structure. The levels are asymmetric, transitive, and require property inheritance. Asymmetry refers to the fact that all boats are vehicles, but a vehicle is not necessarily a boat. The transitive property follows that if a sail boat is a boat, and a boat is a vehicle, it can be inferred that a sail boat is also a vehicle. Subordinate items also inherit all the properties of the level above them. All boats can float and move across the water, therefore all sail boats must as well with the additional property of having a sail. While there are exceptions to these rules, this framework is

the basis for understanding the relation for how we learn and use categories in our everyday lives.

Behaviorally, we know a great deal about how these hierarchical levels affect a range of tasks. In exploring how fast we can verify the category of objects, a basic level superiority was discovered (Rosch et al., 1976). This superiority effect is believed to be caused by a favorable balance existing between object specificity and object distinctiveness (Murphy & Brownell, 1985). Subordinate level objects tend to have very specific features, which often overlap more with similar object categories and therefore lack distinctiveness. Conversely, superordinate objects are highly distinct, but the features are so variable within the category that the category members generally lack specificity. It has also been theorized that different features are more or less important depending on the hierarchical level. Aiding the basic level in visual tasks is that shape information tends to be representative of category membership (Rosch et al., 1976). Alternatively, color information may be more diagnostic when making subordinate level judgments (Hagen, Vuong, Scott, Curran, & Tanaka, 2014).

Computationally, hierarchies have been leveraged to aid the problem of object recognition by computers. This issue is a fundamental cornerstone in the field of artificial intelligence research. Any AI system would have to mirror human's exceptional ability to find, recognize, and use an object according to its categorical properties. Zweig & Weinshall (2007) found that by organizing their object recognition models into a hierarchical structure, recognition rates increased. They also demonstrated that it is possible to learn new object classes, even with just a few small set of objects, by transferring information about neighboring categories. Currently, state of the art object detectors are all convolutional neural networks (Krizhevsky, Sutskever, Hinton, 2012; Hoffman, et al., 2014). The underlying math for these models are

decades old, but have made a resurgence over recent years aided by the cheaper cost of computing power. While they may not explicitly classify objects into a hierarchical structure, the nature of convolving and pooling larger and larger regions over a number of layers is not too dissimilar to the neural networks used by the visual system. Examining the features in each of the layers shows a growing complexity, starting with rough edges and gabors to object parts until eventually objects themselves are reconstructed. Currently, work has begun on a convolutional neural net that is informed by human physiology that attempts to extract the important features from these layers (Yu & Zelinsky, in prep). This is an extension of the Category-Consistent Feature model which is discussed and used in experiment 3.

This dissertation is specifically focused on highlighting and reviewing the current state of understanding categories and expanding on this knowledge in a number of key areas. These include the modelling of categorical representations, their neural correlates, and the behavioral factors involved in using them to perform the everyday task of visual search. My goal was to not only advance these key areas of research, but also to demonstrate that there are still large gaps that need to be filled. Specifically, those gaps exist in the impoverished stimulus sets that researchers have thus far been limited too. Categories are fluid constructs that can encompass countless exemplars. While all of our models of how we learn these categories are restricted to basic visual features or toy examples, our visual categorical tasks are only able to test a small percentage of object or scene exemplars that make up the complex relations of any given category. The purpose of this dissertation is to make measurable steps towards the goals of stimulus enrichment and scalability that we might better understand the characteristics of a broad number of categories with thousands of exemplars.

Experiment One Introduction

One of the most common everyday tasks that uses our hierarchical categorical knowledge is to find something we are looking for. Take a second and try to recall the last thing that you looked for. When you started to search for that object, were you shown a picture preview of exactly what you were looking for, or did you need to rely on retrieving a representation from your memory about previous instances of seeing it or other similar objects? For decades, visual search research has focused on the inorganic task of cueing participants with a picture preview of what to search for. While this has allowed for rigorous control over the basic features that can guide search, it is lacking in its ability to generalize to the search tasks that are by far the most frequent we perform, categorical search. Initially, it was even thought that categorical search was guided weakly or not at all (Vickery, King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004), meaning there was no efficiency in where our eyes moved to localize the target, looking at objects in a serial manner until we stumbled upon the target. Recent years have largely seen a reversal in this thinking, with more and more research elucidating the factors that underlie categorical search (Yang & Zelinsky, 2009; Malcolm & Henderson, 2009; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009).

Whether it is in scenes, search arrays, or virtual environments, with familiar or unfamiliar categories, it is important that we extend our understanding to objects encountered in the visually feature rich world that we live in. Understanding the factors that affect categorical search can aid us in optimizing the search performance of experts and non-experts alike. Whether it is increasing the detection rate of radiologists (Drew, Vo, & Wolfe, 2013) or that of transportation baggage screeners (Meuter & Lacherez, 2016), there are costly consequences for misses in low target frequency situations such as these. Conversely, it is also not always advisable to search

for objects at a subordinate level. When speed is a critical factor, such as being a submarine sonar operator, there is evidence that superordinate level judgements would be the most rapid (Mack, Wong, Gauthier, Tanaka, & Palmeri, 2009). While searching for a lime in a grocery store or navigating a trip through the subway may not have as dire a set of repercussions, these situations are important based on the sheer frequency that they and other similar search situations occur.

Since most any object can be called by many different categorical names, it is important to understand how these varying levels can elicit different performances in identical tasks. For our purposes, and as is common, I will be focusing on the basic, subordinate, and superordinate levels. It is important to note that categories are not confined to simply three levels of categorization, but it is a necessity for controlled experiments to define and make a best attempt at trying to make a stimulus set that is similar in this regard. For example, a basic level category could be 'dog'. A subordinate category is 'golden retriever' and a superordinate one an 'animal'. However, another label that could apply that is still superordinate to 'dog' could be 'mammal'. The distances between these categories is essential to precisely defining the feature space within and between these hierarchical levels. This particular issue is specifically dealt with later in experiment three. For our purposes now though, it has to suffice that it is not the number of subordinate or superordinate labels a basic level category has (or even how you define a basic level category), but the hierarchical relationship as governed by the loose set of rules mentioned previously.

One recent categorical search study specifically looked at the effect that the hierarchical level of the cue had on performance (Maxfield & Zelinsky, 2012). Targets were designated using either a word cue or a picture preview, and participants were asked to search through an

array of six objects consisting of a target or a categorical lure among non-targets from random superordinate categories. Among the categorical cue conditions, guidance, which was measured as the proportion of trials in which the target was initially fixated, was strongest following a subordinate-level target cue (e.g., ‘‘taxi’’). However, target verification, measured as the time from target fixation to the search judgment, was fastest for targets cued at the basic level (e.g., ‘‘car’’), replicating the basic-level superiority effect (Rosch et al., 1976; see also Murphy, 2002) in the context of a search task. This dissociation between search guidance and verification by hierarchical level draws attention to the importance of categories in modulating search behavior, and the need to better understand the factors affecting this modulation.

One limitation to this study however is the static set size of six objects. While the search arrays were designed to allow the relatively strict guidance measure of whether the target was first fixated, we were not able to make any claims as to whether the overall pattern of results would change in the presence of more distractor objects being present. If we think of a category as a conjunction of features, it is the natural extension for visual search research to advance to. A static set size of random distractors means that any change in the signal to noise ratio in the search display is generated by the hierarchical level of the cue. While this has been useful in the past, it is beneficial to extend these findings to a broader range of possible signal to noise values. By adding additional distractors on top of the hierarchical cues we are doing just that. This is an important question to address as it has implications for understanding the characteristics of hierarchical cues and how it affects our search behavior. It is important because when we search for an object in our environment, there will always be a variable number of distracting objects. These objects may share some, none, or all of the visual properties of the target. The initial search epoch is particularly sensitive to this target-distractor similarity. Presented with any

number of possible locations to move the eyes, the visual system has to determine which of the multiple objects is most like the cued category. Alternatively, during the verification epoch, fixations dwell for longer on a single object. This dwell time is likely reflective of the time needed to increase confidence that the target being looked at is enough like other members of the category to also be included in it. Where guidance is a search for a target among distractors, verification is a confirmation of the members that make up the target category. By manipulating the set size in addition to defining these search epochs, it is possible to test the hypothesis that additional distracting features will differentially impact being cued across hierarchical levels.

Based on what we know already about the specificity and distinctiveness of hierarchical levels, by varying the set size of the search display, we are able to explore the resulting behavioral effects. Of particular interest is if the subordinate level remains the fastest in search guidance. While this advantage has been the strongest effect in previous work, it is possible that at higher set sizes the effect is diminished. This would potentially be caused by a decreased search signal amongst the increased noise of multiple distractors leading to more inefficient search behavior. An additional hypothesis is if the basic level cues are more resistant to additional distractors during verification, as their balance between specificity and distinctiveness may afford them some immunity to additional distracting features.

Methods

Participants. Twenty-seven Stony Brook University undergraduates completed the experiment for course credit. All provided informed consent. Participants were native English speakers in an attempt to roughly equate understanding of the categorical cues and the exemplars that belong in each category. They also self-reported having normal or corrected-to-normal visual acuity.

Two participants were removed from any analyses because of accuracy scores < 65% across all trials.

Apparatus. Eye position was sampled at 1000Hz using an Eye Link 1000 (SR Research) eye-tracker with default saccade detection settings. Calibrations were only accepted when the average spatial error was less than 0.5°, and the maximum error was less than 1°. Head position and viewing distance were fixed at 65 cm using a chinrest. The search arrays were presented on a flat-screen LCD monitor at a resolution of 1440x900 pixels. Objects subtended ~2.5° visual angle and category names were drawn in 18-point Tahoma font. Judgements were made using the left and right index finger triggers of a game pad controller.

Stimuli. Images of common objects were obtained from ImageNet (http://www.image-net.org) and various web sources. All images were closely cropped using a rectangular marquee to depict only the object and a minimal amount of background. Because object typicality can affect categorization and search (Murphy & Brownell, 1985, Maxfield et al., 2014), targets were selected to be typical members of their category at the subordinate, basic, and superordinate levels. We did this by having 45 participants complete a preliminary norming task in which 240 images (5 exemplars from each of 48 subordinate categories) were rated for both typicality and image agreement (Snodgrass & Vanderwart, 1980) at each hierarchical level using a 1 (high typicality/image agreement) to 7 (low typicality/image agreement) scale or 0 to indicate that the object did not belong to the category. The three most typical exemplars of each category with no 0 ratings were used as targets in the search task. Figure 1A lists these category names while Figure 1B displays a sample of the exemplars. In total there are 68 categories spanning 3 hierarchical levels; 4 superordinate-level categories, each having 4 basic-level categories, with each of these having 3 subordinate-level categories.

Search Procedure. Participants were shown a category name for 2,500 ms, followed by a fixation cross for 50 ms and then a search array. There were arrays with a set size of 6, 12, and 18, with objects randomly arranged on the screen with no two objects being within ~2° of each other and within 3° of the center of the screen (Figure 2). Each categorical cue was presented once per set size. Counterbalancing across subjects was done so that each participant had half of their trials be target present and the other half target absent, with the counterbalancing making it so this 50-50 present-absent split is equal across set size conditions. On half of the target absent trials, a lure trial was presented to ensure encoding at the cued level (Tanaka & Taylor, 1991). As opposed to the random distractors on the other target absent trials, a lure was an exemplar from the same parent category as the cued target. For example, if the participant was cued with sail boat, a lure trial might display a picture of a cruise ship.

Results

Significant differences in accuracy between cueing conditions were found in both the target-present and target-absent trials ($F(5, 120) = 35.34$, $p < .001$). Post-hoc tests (LSD corrected) on the target-present data revealed that error rates were significantly higher for superordinate cues ($M = .77$, $SD = .09$; p's $< .001$) than basic ($M = .87$, $SD = .07$) and subordinate ($M = .86$, $SD = .07$) conditions. This miss rate is approximately the same as previous experiments and stems from miscategorizations during a speeded task compared to the slower and more deliberate norming task. Target-absent trials with a subordinate cue yielded significantly higher errors (p's $< .001$). This is accounted for by the increased difficulty and uncertainty presented by the lures at this level (Non-Lure Accuracy, $M = .96$, $SD = .03$, Lure Accuracy, $M = .79$, $SD = .10$). There were also significant differences between set size conditions ($F(5, 120) = 30.38$, $p < .001$). Post-hoc tests revealed accuracy was highest for target-present 6-object arrays, followed by 12- and

18-objects (p's < .001). On target-absent trials there were no differences between the three set size conditions (p's > .49). The additional distractors had the desired effect of increasing the difficulty of the search task. Only correct trials were included in the subsequent analyses.

Consistent with previous studies (Castelhano et al., 2008; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009; Yu, Maxfield, & Zelinsky, 2016), we divided our analysis of search performance into guidance and verification epochs. Search guidance was defined as the time from the onset of the search display until the participant fixated the target (time-to-target). Verification was defined as the time between when a participant first looked at the target and when the target-present/absent button response was registered.

Analyses of search guidance for target-present data revealed significant differences in time-to-target across both set sizes ($F(2, 48) = 326.01$, $p < .001$) and hierarchical level ($F(2, 48) = 37.89$, $p < .001$). As set size increased, so too did the average time-to-target (p's < .001). When examining the role of hierarchical cue on time-to-target, a significant linear trend was found which showed that on trials with subordinate cues, the target was fixated faster on average than basic-level cues, followed by superordinate cues (p's < .001). These differences remained when taking both set size and hierarchical level into account ($F(8, 192) = 94.97$, $p < .001$; Figure 3A). The subordinate level advantage in time to target held for set sizes of six and twelve, however with a set size of 18, there was no significant difference between subordinate and basic level cues ($p = .378$). Initial saccade latency did not reliably differ between cueing conditions or set size (p's > .511), suggesting that the differences in time-to-target were not due to a speed-accuracy trade-off. The time-to-target guidance measure generally supports the previous reports showing an increase in search efficiency with greater cue specificity (Schmidt & Zelinsky, 2009; Maxfield & Zelinsky, 2012). However, the set size manipulation revealed an interesting finding.

There may be an upper limit to the advantages afforded by cue specificity when more distracting objects are present in the display.

Turning to target verification, there were again significant differences across the set size and cueing conditions ($F(8, 192) = 38.32$, $p < .001$, Figure 3B). Post-hoc tests revealed that the superordinate level was significantly slower than both the subordinate and basic levels across all set size conditions (p's < .002). Most surprisingly though was that there was no basic-level advantage observed at any of the set size conditions. Instead, there was simply no significant differences between subordinate-level and basic-level verification times at any of the set size conditions ($p > .100$). This is an unexpected deviation from previous research that used the same stimuli and the same number of distractors from a previous study.

Discussion

The present study sought to explore the effect of set size on a categorical search task. By varying the number of distractors and observing the behavioral results, we can extend our knowledge of how various hierarchical cueing conditions impact search performance. Overall, we saw the expected pattern in the overall reaction times increasing with a higher set size. This pattern was also consistent when broken down by how participants were able to fixate on the target, again demonstrating a subordinate level advantage where more specific cues lead to more efficient guidance. This advantage was only diminished at the largest set size when basic level cues were just as fast. This particular result suggests that there may be a point at which once there are enough distracting objects in the visual field or those objects are sufficiently close together, the subordinate level advantage ceases to be. At this point, the advantages lent by the specific features in building a target representation are not enough to offset the increase in the

number of distracting objects. Visual crowding can impact object recognition at any number of stages during visual analysis (see Whitney & Levi, 2011). The specific type of crowding present here though is likely an interference of specific features (Levi, 2008).

In terms of verification times, we did not see the expected basic-level advantage for a set size of six objects, let alone as set size increased. This may suggest that the basic-level advantage is a much more fragile effect than years of category verification experiments would suggest. If the effect was not so fragile, subordinate level cues would not be verified as fast or faster across all set sizes. As for the specific lack of an advantage at six objects, it is impossible to make conclusive statements from a null result. However, I think there are two large differences between this experiment and our previously conducted categorical search task. First is the arrangement of the objects. Whereas previously they were arranged in a circle, where objects were always flanked by 2 distractors in parafoveal view, this experiment was arranged in random locations (given the constraint that they did not appear too close to the center or too close to one another). Even though the number of objects was the same, the average inter-object distance was vastly greater in the present experiment (~1.5° versus ~2-10°). This may have possibly led participants to explore other possible alternatives less once they had landed on a target, giving way to the non-significant differences between subordinate and basic-level cues. The other main difference was the set size manipulation itself. The basic-level advantage is an effect predicated on the presence of lures on target-absent trials. So we know that the effect can be mediated by certain task demands. It may be the case that just by simply manipulating the set size, participants would change their criteria for making the verification decision.

To say that in the real world, the number of distracting objects present when completing a search task is unconstrained would be a gross understatement. By just selecting three different

set size conditions, we were able to find small nuances in how people respond at various hierarchical levels. These results lend strength to the argument that there is still a great deal to learn about how the hierarchical level of the cue provided can affect our behavior. The prevailing logic from the visual search literature was predominant. That is, the more specific the cue, the more specific our target template will be. It is that match between the template and the search display that usually leads to the best performance. However, there seems to be a plateau in performance where both subordinate and basic level cues are indistinguishable. It is possible that adding more lure trials would result in increases in subordinate level cue response times. A future experiment may look at manipulating the proportion of target-absent trials to give basic-level cues as much of a chance of an advantage during verification as possible.

Experiment Two Introduction

Visual search is a complex task likely involving a complex interaction of serial and parallel processes working in conjunction to achieve a goal. In a recent review, Eimer (2014) proposed a four-stage model of selective attention in visual search. These stages are useful in understanding the framework for just how complicated a task visual search is. Each stage is associated with a cognitive function and a set of neural processes associated with this function. Preparation is the first stage in which a target template, a representation of the target object, is loaded into working memory. This is associated with a goal-selective pattern of neural activation in visual and/or pre-frontal areas. Guidance is the next stage where there is a massive parallel accumulation of task-relevant information. Whatever relevant features are being used to guide search, orientation, color, or motion for example, neural activity will spike in the brain region associated with it while regions not relevant will be inhibited (Martinez-Trujillo & Treue, 2004; Cohen & Maunsell, 2011). This is followed by the selection stage where candidate target

objects are localized to specific locations. It is here that we find a neural correlate with a spatially selective enhancement of visual responses to these possible target objects. One way that we can measure this enhancement is using electroencephalography (EEG). Specifically, an N2pc component is observed. The N2pc is a temporally precise event-related potential that provides an index of this selection process for a target among distractors (see Woodman & Luck, 1999; Eimer, 1996; Luck & Hillyard, 1994). It is found by recording the neural activity over the parietal and occipital lobes that is ipsilateral to the target object and subtracting it from the contralateral activity in the same area. It is normally expressed 180-300 msec after stimulus onset (Hopf et al., 2000). This component will be one of the points of focus in the current experiment. For completions sake though, the final stage is that of identification. This is when selected objects in working memory are maintained for integration and discrimination. When this happens we can observe a sustained activation of object representations from recurrent feedback loops (Luck & Vogel, 2013).

Since we are dealing with processes that happen so rapidly and are difficult to pinpoint when one stops and another begins, especially along with all the actual motor programming that is happening simultaneously, EEG becomes a very useful tool. There are a few ERPs that are of particular interest, namely the N1, N2pc, P3, and contralateral delay activity (CDA).

The closest comparison to the present work was an examination of subordinate, basic, and superordinate cues on category verification performance (Tanaka, Luu, Weisbrod, & Kiefer, 1999). In it, they found an N1, characterized by an enhanced negative deflection in left posterior channels on subordinate trials compared to the basic level. This was interpreted as an indicator of increased visual analysis on account of the more specific features at the subordinate level. Additionally, they reported increased frontal negativity for the superordinate condition relative to

the basic level. This result was attributed to increased semantic processing of superordinate categorization. In contrast to their study, our study had multiple objects present in the search display just outside of foveal vision. The N1 can appear in different cortical locations depending on the spatial location of the stimuli (Wascher, Hoffman, Sanger, & Grosjean, 2009). So while we might expect to see an N1 present in our study, it may not be restricted to just the left posterior channels.

A number of studies have been conducted in an attempt to understand the neural components of selecting real world objects in a visual search task when cued with a categorical label. The main conclusions reflect the behavioral research, that categories are able to efficiently guide search (Wu, Scerif, Aslin, Smith, Nako, & Eimer, 2013; Nako, Wu, Eimer, 2014). One study found that the N2pc was attenuated and delayed when searching for a target cued by a category compared a pictorial preview (Nako, Smith, Eimer, 2015). This fits with previous findings on the advantages of having a precise target template provided by knowing what the target object will exactly look like (Schmidt & Zelisnky, 2009; Maxfield & Zelinsky, 2012). Interestingly though, they also found that the imageability of the objects also affected the N2pc. The more the cue had the ability to activate a target-matching visual representation the higher and earlier the peak N2pc amplitude was found.

The P3 component is characterized by a positive deflection in centro-parietal brains areas in response to categorization of the presented stimuli. This effect has been well researched with experiments covering the categorization of auditory stimuli (Mecklinger & Ullsperger, 1993), its role in childhood development (Batty & Taylor, 2002), and novel stimuli (Polich & Comerchero, 2003). Examination of this component in regards to differences in visual stimuli has yielded evidence that there is differential ERP activity for photographic objects that were degraded or

were defined on semantically different levels of categorization (Johnson & Olshausen, 2003). Their comparison of targets cued at either a basic or superordinate level revealed that the average amplitude differences was nearly twice as large in the basic condition while superordinate cues had shallower amplitudes with a later onset. This characterization largely follows previous research on the time course of superordinate level categorization, though there was a roughly 100ms difference in what each study concluded was the precise time this categorization occurred (Tanaka, Luu, Wisbrod, Kiefer, 1999). The present study is well placed to clarify the findings of targets cued at the superordinate level as well as add the additional specificity of subordinate level cues.

The first step to completing a search task is knowing just what you are looking for. This preparation, the loading of target relevant features into a template, is vital to success (Desimone & Duncan, 1995). The result of a poor fit between the features of the template and what the target actually looks like has been demonstrated by numerous methods including rotation, color, scale, and level of categorization. In the time between the cue and onset of the search display, as well as during the search task itself, neural firing rates of target relevant features are enhanced in the inferotemporal cortex (Chelazzi, Miller, & Duncanf, 1993). This template is thought to make demands on the visual working memory (WM) system (Duncan & Humphreys, 1989; Bundesen, 1990). In a recent study, CDA, commonly used to measure WM load, was investigated during a variety of search tasks that varied in difficulty (Luria & Vogel, 2011). They found that as search difficulty increased, so too did the role of WM. This suggests that to the extent that a search based on a superordinate cue is more difficult than a subordinate one (likely stemming from a sparser target template), we might also expect to see differences in CDA. Alternatively, given

that we know subordinate cues contain more specific features, an increase in CDA could be interpreted as loading more of these useful features into WM.

There have been a few studies that have explored the topic of categorical search, but none that detail the effects of manipulating the hierarchical level of the cue. The purpose of the proposed study is to examine this question. Specifically, do all hierarchical levels elicit the same neural components? Are there any differences in the amplitude or onset of these components? Does imageability also vary by hierarchical level, possibly accounting for any further differences in the N2pc that we find between levels? Do cues that vary by hierarchical level place more or less of a burden on WM? Will the semantic differences of the cues themselves affect the categorization decisions that must be made to complete the task?

Methods

Participants. Twenty-two participants from Stony Brook University whose native language is English and have self-reported normal or corrected-to-normal vision participated in this study. All participants gave informed consent and were compensated with course credit. Six participants had a >35% of their trials rejected from eye movements or blinks and were therefore not included in any analyses. All data reported is from the remaining sixteen participants.

Apparatus. Eye tracking data was collected in the same manner as experiment one. EEG data was collected from a 64 channel Biosemi system, with scalp electrodes at standard positions using the extended 10/20 system. All recordings were referenced to two mastoid channels. Horizontal eye movements were recorded from electrodes 1cm to the left and right of the lateral canthi in order to measure horizontal eye movements. Vertical eye movements were detected by an electrode placed below the right eye. Trials that had eye movements, as identified by

horizontal electrooculogram (EOG) exceeding +-30 V, vertical EOG exceeding +- 60 V, and all other channels exceeding +- 80 V, were removed. We segmented the continuous EEG recording -200 to 1000 ms relative to the search array onset. Waveforms were baseline-corrected for the 200 ms interval prior to the presentation of the search display. Grand averages were obtained by averaging across the subject averages for each experimental condition. Difference waves for analyzing the N2pc and CDA were computed by subtracting the activity on the hemisphere contralateral to the target stimuli spatial location from the activity at the corresponding ipsilateral recording sites.

Stimuli. This experiment used the same stimulus set as Yu, Maxfield, & Zelinsky (2016) and experiment one. It includes three exemplars per subordinate category that are cropped images from the ImageNet database. These images have been normed and found to be typical exemplars of their categories and have high image agreement.

Procedure. Participants first completed a set of forty practice trials in order to assess their ability to maintain central fixation while completing the task. If an eye movement or blink is detected on greater than 40% of the practice trials, the participant was told they are not eligible to complete the study and received partial credit. If the participant was eligible, they underwent the preparation necessary for the EEG recording.

The experimental trials consisted of each categorical cue being presented twice at each hierarchical level, once as target present, and once as target absent. On half of the target absent trials, a lure was presented. The categorical cue appeared for 2,000ms followed by an 'x' in the center of the screen for 1,000ms, after which the search display appeared. The search display contained four objects arranged in a circle with an eccentricity of 3°. The search display

remained on the screen until the participant chose if the target is present or not by triggering a button on a gamepad controller (Figure 4). Any eye movement exceeding 1° from the beginning of the categorical cue until a button is pressed displayed a message informing the participant to not blink or look away until the trial is over. These interrupted trials were removed from any analyses.

Results

Behaviorally, we found no significant differences between accuracy on target present trials when comparing conditions, $F(2,30) = .167$, $p = .847$. Partialling the accuracy data between trials in which there was an eye movement or a blink which terminated the trial compared to incorrect button responses revealed that on average 14.3% of trials were removed for eye movements and 24.5% were target misses. On target-absent trials, catch trials had a significantly higher error rate than non-catch trials, $t(15) = 7.15$, $p = .001$. An analysis of reaction times on correct target-present trials revealed significant differences, $F(2,28) = 4.59$, $p = .05$. Post-hoc tests revealed that while there was no reliable differences between subordinate and basic response times ($p = .682$), superordinate trials were responded to slower on average (p's $< .05$; Figure 5).

A number of ERP components were analyzed across various time frames relative to the search display onset. Recording sites selected for comparison were chosen by locations where the component typically appears based on previous literature in addition to computing the peak voltage amplitudes across a range of our recording sites. The time window selected for analysis was also informed by the same methods. All ANOVAs for this experiment were performed with the Greenhouse-Geisser correction to account for the sphericity departures that are common when analyzing ERP components (Geisser & Greenhouse, 1959).

Evidence for an N1 component was examined by analyzing the averaged waveform from a time segment 150 to 200 ms after the onset of the search display. We found no significant differences between mean amplitudes for any condition in left posterior channels (averaged activity from CP1, CP3, CP5, TP7, P7, P5, P3, P1, PO3, PO7, O1), right posterior channels (CP2, CP4, CP6, TP8, P2, P4, P6, P8, PO4, PO8, O2), or a combination of all posterior channels, F's < .878, p's > .386. A time segment ranging from 300 ms to 700 ms was analyzed for a potential P3 component. Mean amplitude averaged between midline recording channels (Fpz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz) revealed no significant differences between conditions, $F(1.715, 25.73) = .684$, $p = .492$. An investigation of any possible N2pc effects was computed in the 250 to 300 ms time range. There were no significant differences when mean amplitudes were calculated from parieto-occipital channels (O1/O2, PO7/PO8, P9/P10), $F(1.8, 26.99) = .09$, $p = .896$. This was also the case when analyzing these pairs of channels individually. Analyses on CDA were run from a time window of 350 to 600 ms. Again, no significant differences in mean amplitudes were found between cueing conditions at the channel pairs of O1/O2, PO7/PO8, PO3/PO4, F's > 2.11, p's > .14. All analyses were performed again within +/- 50 ms of their respective time windows, but no significant effects emerged between conditions. A visualization of the lateralized activity between 150 – 500 ms can be found in Figure 6.

Through further investigation of the waveforms for all of the channels, we noticed a large negative peak around 200 ms after search onset in right posterior channels. This negativity was not dependent on whether the target was present on the left or the right side of the screen. Eventually this negativity spread to the left posterior channels for all conditions between 300 and 400 ms, before spreading further into frontal regions near the trials conclusion. Such unilateral

activity could possibly be overwhelming any chance of finding a lateralized effect in that region, such as the N2pc and CDA.

Discussion

The purpose of this study was to attempt to correlate the behavioral differences elicited by hierarchical categorical cues with neural activation in the form of ERP components. Participants were asked to respond to parafoveally presented images of real world objects and make a present or absent decision while having their gaze remain centrally fixated throughout the task. Unfortunately, while there were slight differences in a number of key areas, no significant differences emerged in the EEG data between conditions. Potential reasons for this are discussed below.

In terms of the behavioral data, the null result between subordinate and basic bears some consideration while superordinate cues being slowest was as expected. While this task shares the same hierarchical structure and images as other studies we have conducted, the layout of the objects means that calling this a search task is perhaps taxing the definition somewhat. The closeness of the objects to central fixation puts this somewhere between a covert search task and a category verification task with multiple distractors present. Since gaze had to remain centrally fixated, it is impossible to infer whether these changes helped basic level cues speed up because the verification process could begin almost immediately once the search display appeared, or if the subordinate condition was slowed down since overt guidance to the target was not necessary. This result could potentially inform changes to a follow-up study though. For example, by systematically manipulating the eccentricity of the objects from central fixation, it should be possible to determine how much of a benefit the basic level was afforded by being so near the

central fixation point. The target-absent trials, specifically the ones with lures, are there to ensure encoding at the level that was cued. If participants are doing that anyway, it would help increase the power of the target-present analyses if there were fewer target-absent trials and more target-present. In any case, the distinction between the two and the superordinate condition would hopefully still be enough to contrast neural signals with the superordinate condition. Unfortunately though, this was not the case.

The lack of significant differences in the EEG data between cueing conditions is somewhat perplexing. There is enough evidence for neural differences emerging from hierarchical cues from similar experiments to expect that we would also find differences in the number of locations and time windows that we investigated. It is possible that there were N2pc and CDA differences between conditions, but they were overlapped by a much stronger negativity in the parietal region to be detected and separated. If the study were to be rerun there could be a number of potential changes to possibly receive some positive results. Cueing conditions could be blocked in an attempt to minimize any variance in the waveforms from continually switching the representation used between hierarchical levels. Alternatively, power could be increased by only examining two of the three hierarchical levels we chose. While this would somewhat diminish the theoretical claims that we would be able to make, it may provide enough power to the analyses that the small differences observed would be significant.

Experiment Three Introduction

One of the biggest obstacles in using real world objects as stimuli is the sacrifice of feature control. To avoid this, simple objects or readily nameable features have been the preference in generating models of how we learn new categories (Shephard, Hovland, & Jenkins, 1961; Lee &

Navarro, 2002). However, we need to ensure that the principles uncovered in past research generalize to the much more complex features of real world categories. The problem is then, what makes two categories different or similar to one another? It's certainly the case that categories are fluid constructs. We constantly learn new ones or update existing ones with the vast amount of exemplars we encounter every day. As mentioned earlier, hierarchical levels are simply rules of thumb to describe how categories are related to one another. Using the hierarchy found in the preceding experiments as an example, how can we quantify the similarity between sibling categories that share the same parent? Are 'oreo' and 'sugar cookie' more or less similar than a 'biplane' and a 'passenger airliner'? I will discuss two potential methods for making just this sort of judgment and why considering this question is an important one.

This question of similarity is not new to the categorization field. Multiple models and hundreds of papers have attempted to describe or quantify it in an appreciable manner to varying degrees of success (Goldstone, 1994, Goldstone & Steyvers, 2001; Nosofsky, 1986; Nosofsky & Palmeri, 1997). One tool that has proven itself useful though for many years has been that of multidimensional scaling (see Kruskal & Wish, 1978; Hout, Papesh, & Goldinger, 2012). Multidimensional scaling (MDS) is a set of statistical procedures that takes input in the form of participants, often making pairwise judgements of similarity between objects on a likert-type scale. The output of the MDS analysis is a map in which all of the objects are mapped to a coordinate in an n-dimensional space, where a closer proximity in coordinate space indicates a closer relationship in terms of similarity. Importantly, MDS is essentially a means of exploratory data analysis and dimension reduction.

With a tool that has been around so long, there is of course no one set way of running an MDS experiment or analyzing it (Jaworska & Chupetlovska-Anastasova, 2009). One downside

to MDS has always been the time needed to collect enough data from participants making pairwise judgments between every single object you are interested in. As the number of objects you are interested in quantifying similarity goes up, the time needed to collect all of the data also rises exponentially. One recently created method that can avoid this problem is known as Spatial arrangement method (SpAM; Hout, Goldinger, Ferguson, 2013). This method allows for participants to see multiple objects on the screen at the same time and spatially arrange them to make their similarity judgment, the greater the distance the more dissimilar the objects are.

Another more recent way to quantify the similarity between categories and the exemplars that constitute them is the Category-Consistent Feature (CCF) model (Yu, Maxfield, & Zelinsky, 2016). The CCF model is also by its nature one of dimensionality reduction, but in its case the input is derived from extracting the local visual features using a computer vision algorithm, SIFT (Lowe, 2004) along with a color histogram (van de Weijer & Schmid, 2006), and then clustering these features using a bag-of-words model. From this point, each object is now represented by a 1064 bin histogram of features. Each histogram is then averaged together to create a categorical representation, with each feature now having a mean and variance associated with it. Borrowing tools from signal-detection theory, two passes are done on these categorical histograms, pruning away the unimportant features with low means and/or high variance. What is left are consistent features that appear often in the category, what the model calls category-consistent features. The model has been successful at demonstrating the behavioral pattern seen in hierarchical search tasks. A by-product of it is that we are now able to quantify the distances between any two given categories of objects by measuring the distance between the features of those categories. This process is completely independent of any human judgment ratings. This obviously leads to the

advantage of being able to collect nearly an endless hierarchy of images, but it is as of yet untested on more complicated categorical relationships where context may play a key role.

There are inherent strengths and weaknesses to both the MDS and CCF approaches, a few of which have been mentioned previously. MDS suffers from the practical limitations of collecting data from human observers. Participants can get bored or distracted if shown too many images, which could lead to artificial effects in the results. However, if it is human categorization that is of primary interest, then MDS provides the best means to collect as close to a groundtruth as possible. It also allows for semantic and regional distinctions, reflecting the fact that while every individual person has a unique representation of any given category, these representations can have complex relationships and are most likely similar to those living geographically close to us. Alternatively, CCF boasts a much higher ceiling in terms of the number of categories and exemplars to consider. While these categories seem to map on well to the factors driving human behavior, more testing needs to be done in different modalities to test the robustness of these claims.

MDS and CCF both allow for the quantitative measurement of similarity between categories of objects. While it can be useful to compare and contrast the inherent strengths and weaknesses to each approach based solely on their theoretical and practical merits, an even stronger test is to judge their performance against a common behavioral experiment whose well replicated results are thought to be based on the similarity between object categories. Category verification is a simple task where participants are typically cued with a category name at the subordinate, basic, or superordinate level. After a brief delay, an image appears and the participant has to make a true or false judgment as to whether that image belongs in the cued category. The response time is seen as an approximation of matching the features of the image

to the cued features that comprise the cued category. It was this very task that uncovered the first of many advantages the basic level seems to have (Rosch, 1976), so it seems an appropriate test for comparing our two methods of interest in this experiment. If both MDS and the CCF model results are able to correlate with the category verification results, the question becomes which is the stronger predictor, which can be answered statistically. Both methods though would have demonstrated their ability to predict performance on another behavioral task. If one or the other should fail to correlate significantly, it is reasonable to conclude that for this specific task, one is a more appropriate tool when predicting category verification decisions. If both methods fail, it would by no means undercut the merit of both approaches to understanding performance in other tasks.

Methods

MDS. The Multidimensional scaling data collection was a collaborative effort between Stony Brook University and Dr. Michael Hout and Arryn Robbins at New Mexico State University. Using the same 144 object stimulus set as experiment one, participants completed a spatial arrangement task wherein they click and drag the image exemplars from each category into a bounding box (Figure 7). Sixty-two participants from NMSU and forty-nine participants from SBU provided informed consent and completed the task for course credit. They were instructed to arrange the images based on how similar they are to each other, with a closer distance in pixels indicating higher similarity. Twenty-five randomly sampled objects were present on a given trial and there were a total of twenty trials. While this means that not every image appeared with every other possible image on a given trial, across trials, based on previous research we will have enough comparisons to other categories to be confident in the resulting coordinates in MDS space. Based on simulations, this will yield around 65 observations per

object. Participants were recruited from Stony Brook University and New Mexico State University. This allowed for a check to control for regional differences. The experiment was run in E-prime on a widescreen LCD monitor at a resolution of 1028x768. Objects subtended ~2.5 degrees of visual space.

CCF. The methods for how the data for the model were computed is identical to the instantiation from Yu, Maxfield, & Zelinsky (2016). The core principle of the model lies in dimensionality reduction. It is able to detect the visual features of categories of objects, extract the important ones that appear frequently and consistently, and then calculate the distances both between and within categories.

The Category-Consistent Feature model was first trained on 100 exemplars for each of the 48 subordinate-level categories (4,800 images in total), with each exemplar being an image patch closely cropped around the depicted object. The basic- and superordinate-level categories were built by combining the subordinate "children" exemplars under the "parent" categories. For instance, the basic level car category had 300 exemplars consisting of 100 police cars, 100 taxis, and 100 race cars. The vehicle superordinate category had 1,200 exemplars, made up of 300 exemplars each from the car, boat, truck, and plane children.

SIFT (Lowe, 2004) and color features were then extracted from all of the exemplars. These local features were then clustered and transformed using a Bag-of-Words method to create a 1,064 feature histogram for every image. By averaging these histograms, a categorical representation can be created. Each bin of this categorical representation contained information on how often each cluster of features appears on average (mean bin frequency) and how consistently it appears across objects in the category (bin standard deviation). By simply

dividing these two measures, each bin can be weighted by how important it is to the category. Bins with higher importance scores were kept in the histogram while the empty or low importance bins were pruned away. The resulting histogram is called a Category-Consistent Feature histogram (Figure 8). These CCF histograms can be roughly equated to the centroid of the categories being calculated by MDS. Using these CCF histograms, it is possible to compare similarity, by comparing the distance between the CCFs of the target category and the features of the exemplars composing the target's categorical siblings, which are defined as categories sharing the same parent (one level up in the category hierarchy). The computation for this involved taking the mean over the hundreds of distances between a categories CCF histogram and the BoW histogram of its neighboring categories. This measure of sibling distance is what gives us the ability to capture similarity ratings without any input from human observers that is the focus of using the CCF model in the present study.

Category Verification. Thirty-five participants from Stony Brook University completed the category verification task. All gave informed consent and participated for course credit. One subject was removed from all analyses for not exceeding 60% accuracy. To accommodate the large stimulus set, the exemplars for the target category were counterbalanced across participants, with each participant only seeing one of the three target exemplars.

Stimuli were presented centrally on a flat-screen CRT monitor at a resolution of 1024 × 768 pixels. Head position and viewing distance were fixed at 60 cm using a chinrest. Objects subtended ~2.5° visual angle and category names were drawn in 18-point Tahoma font. Judgments were made by pressing the left and right index finger triggers of a game pad controller; while trials were initiated with a button operated by the right thumb.

Participants were shown a category name for 2,500 ms, followed by a fixation cross for 50 ms and then an object. They were instructed to respond whether the object belongs to the named category as quickly as possible without sacrificing accuracy. Over the course of the experiment, one target exemplar was cued twice at each hierarchical level. Half of these trials were target present and half were target absent. On half of the target absent trials, a lure appeared.

Results

MDS. As anticipated, the number of times each possible pair of stimuli were randomly selected to co-occur in a SpAM trial was on average sixty-five. Similarity ratings were subjected to an MDS analysis, along with scale- and participant-matched random Monte Carlo simulations. These simulations served as a test of the fit of the MDS space. Comparing the stress of the simulations to that of the actual true data collected serves as a check of the goodness of fit of the dimensionally reduced MDS space. This stress check ensures we are adding approximately the right amount of dimensions to account for the input data, but not more than is necessary which would lead to overfitting. Importantly, we not only found that the simulated data had less stress than the true data ($p < .05$), the true data also successfully recovered the subordinate-, basic-, and superordinate-level category clusters within the stimuli. Post-hoc tests revealed that in the three-dimensional MDS space, average distance for each exemplar to its subordinate siblings was smallest, followed by distances to basic-level and superordinate siblings (p's $< .05$). No significant differences were found between categories for simulated data. Category centroids were also computed at each level of the hierarchy, representing the average X,Y,Z coordinates for a group of category members. There was a significant difference between the average Euclidean distance from each exemplar to its categorical centroid ($F(2, 6) = 5.86$, $p = .039$).

Again, subordinate exemplars were on average closer to their centroid, followed by basic level distances, with superordinate being on average the furthest apart (p's < .45). It is these averaged Euclidean distances that were used for the comparison to the category verification task (Figure 9).

CCF. There was a significant difference between average sibling distance between cueing conditions, $F(2,4) = 140.97$, $p = .001$. Sibling distance was smallest for subordinate exemplars, followed by the basic level, with the superordinate level having the highest (p's < .026). This pattern in the mean data is present in the MDS Euclidean distance and CCF sibling distance measures. This can be taken as an indication that there is agreement between both of these measures on the coarse shape of similarity relationships between our categories.

Category Verification. There were no significant differences in reaction times between counterbalanced conditions (p > .05). All further analyses therefore collapsed across counterbalance. There were significant differences found in both target-present and target-absent accuracy (*F*'s > 19.02, *p*'s < .001). Post-hoc tests revealed that on target-present trials, participants were less accurate on superordinate trials (M = 88.2%, SD = 8.1%) than basic (M = 96.3%, SD = 3.0%) and subordinate (M = 96.2%, SD = 2.8%). During the norming of the stimulus set none of the target objects were miscategorized which would have resulted in their exclusion. However, the increased errors at the superordinate level are typical in tasks that require a speeded response such as this one. On target-absent trials, there were significantly more errors on subordinate trials (M = 85.6, SD = 7.2%), specifically those which contained a lure (p < .05). These results are also in line with our previous findings.

For correct, target-present trials, there were significant differences between cueing conditions $F(2,68) = 4.73$, $p = .012$. While there was no difference between subordinate and basic cues, superordinate cues took significantly longer to respond to ($p < .001$; Figure 10). There was no basic-level advantage, contrary to apriori predictions.

Comparison. When correlating the MDS results with the category verification task, the Euclidean distance of the target exemplar was compared to the centroid of the cued category. This is similar to the CCF method where the targets chi-squared sibling distance was measured using exemplars from the cued category. Each of the analyses were broken down by hierarchical level. None of the correlations of the MDS results to the category verification target-present reaction times at either the subordinate $r(144) = .007$, $p = .936$, basic $r(144) = .132$, $p = .115$, or superordinate levels, $r(144) = .129$, $p = .124$, were significant. Likewise, sibling distance was not able to account for the category verification response times at any level (p's $> .302$). A number of other measures were generated from the CCF data in an attempt to capture the behavioral pattern, including the chi-squared distance from the target histogram to the CCF histogram, sibling distances between larger portions of the hierarchy, and a multiplication of the two, all to no avail.

Discussion

The present study attempted to compare the ability of two techniques that compute the similarity of photorealistic objects. MDS uses a spatial arrangement task completed by human participants to calculate similarity, while the CCF model compares histograms of visual features extracted computationally. Neither of these methods though were able to correlate with reaction times in the category verification task despite having similar mean patterns overall.

The results from the category verification task varied from apriori predictions. The lack of a basic level advantage, especially when one was found in the verification epoch of the search task from Yu, Maxfield, Zelinsky (2016), is surprising. Given this, it is worth comparing what differences could account for this unexpected pattern. The search and category verification tasks shared the same target and distractor stimuli and the same categorical cues. Obviously the largest difference was that only one object was presented during the category verification task. It is possible that this contributed in part to the basic-level advantage seen in the search task. Since lures at the subordinate level were more distracting, it forced subordinate verification times higher than the basic level. Whereas in the verification task, the lures still caused a slowdown in reaction time, but not enough for the basic level to emerge at the fastest. Another important factor is a detail of the stimuli that was used. All of the objects were selected from Google and ImageNet and had complex backgrounds. While these backgrounds were minimalized by closely cropping around the target, all of the objects are technically rectangular. This could diminish the usefulness of using shape features for object categorization, which previous research has shown to be one of, if not the most important feature in aiding categorization at the basic level (Rosch et al., 1976).

While it is difficult to interpret what may have accounted for the lack of correlations for both the CCF model and MDS, it is still worthwhile to investigate why this may have occurred. The mean response times, Euclidean distances, and chi-squared distances all followed the general pattern of subordinate and basic level categories being equal or near equal and superordinate targets being the least similar to their categorical centroid. This suggests that there may not have been sufficient variance in either the reaction time data or the similarity metrics to map on well to each other. Additional experiments that focused on increasing the variance from

these measures would need to be conducted in order to parse these possibilities apart. One potential way to encourage a basic level advantage would be to increase the number of lure trials. This would hypothetically increase the reaction times on target present subordinate trials, similar to what was demonstrated in the search task.

Experiment Four Introduction

There is a rich history of quantitative models that attempt to explain visual search behavior (Najemnik & Geisler, 2005; Wolfe, 1994; Zelinsky, 2008). Some of these though are based around the assumption of a very simple target whose features are known, if not explicitly previewed before a search task. Two models that have attempted to extend past this limitation are that of the categorical target acquisition model (cTAM; Zelinsky, Adeli, Peng, Samaras, 2013) and MASC (Adeli, Vitu, Zelinsky, 2015). Both of these models used computer vision methods to train a classifier on a large set of images to learn the likely features of a target category. These features were then used to attempt to localize a categorically defined target in a search display.

The main difference between these models and the CCF model is the dimensionality reduction process that takes place. Whereas cTAM and MASC use the entire feature histogram for a trained object, the CCF model prunes away the unimportant features, leaving those thought to be important in driving search performance. We found that the number of CCFs, those remaining important features for a given category, was able to predict the overall mean pattern including the subordinate level advantage in search guidance. However, an explicit test of the CCF's ability to make finer grained predictions has yet to be conducted. While it has been shown to do well in predicting the overall trends in a hierarchical search task, it has not made

any predictions about search performance on individual trials. The current experiment will attempt to elevate the CCF model from a useful tool in scaling up our ability to characterize the differences and similarities between categories to that of a model of categorical search.

In order to qualify as a model of search, predictions will have to be generated on the much finer grained scale of individual trials. This involves generating a target map that will compare the CCF histogram of the cued category to that of all of the objects in the search display. Such a demonstration would lend strength to the claims that these category consistent features are crucial in driving search performance.

Methods

The CCF model was run on the same training dataset and test images as in experiment three and Yu, Maxfield, & Zelinsky (2016). In previous work, time to target was correlated with the number of CCFs for a given category. However, since we do not know the CCFs for the random set of distractor objects we had to alter the measurement we derive from the model. The chi-squared distance between the cued CCF histogram and the corresponding feature bins of the objects in the search display serve as a viable alternative for measuring guidance (Figure 11). It does not require us to know the CCFs of the distractor objects and will also give us a measurement of how closely the features of the object match the cued category. In order to account for the distractors in the display, the distances of all of the distractors chi-squared distances were averaged together and then subtracted from the targets distance. We are calling this a target distinctiveness score (TDS)[1], as it quantifies the difference of the target to the cue as well as the target and the surrounding distractors. With a value associated with each object, this

[1] It should be noted that this was just one possible way to calculate the effect of distractor distances on the TDS. Another viable method is to only subtract the distractor with the smallest distance (most like the cued category). A more detailed summary of these methods can be found in the computational modeling section.

can serve as a rudimentary target map for how attention could be deployed to complete the task (Figure 12). Multiplying this measure with sibling distance, or the proportion of overlapping CCF bins that a category shared with the remaining categories, was used as the correlate to target verification. Unlike the measure of guidance which took the distractors into consideration, the verification measure was driven solely by the other exemplars within the categorical hierarchy. This is in line with statements made previously in experiment 1 suggesting guidance can be thought of as a search for a target between objects where verification is a confirmation search among the categories exemplars. Each trial was now associated with a TDS and TDS*Sibling Distance score which could be used to predict guidance and verification times respectively. Aside from these necessary modifications, the CCF model was run as previously published with the code publicly available (https://github.com/cxy7452/bow-ccf). These new measures should not only allow us to model the overall behavioral trends, but to also quantify the important factor of target distractor distinctiveness.

Results

Target Present Guidance. As a first check, I averaged the TDS for each of the three counterbalance conditions. This check was done for two reasons. The first was to ensure we began with a similar trend to the behavioral data. If we found anything but the significant linear trend between cueing conditions, no conclusions could be drawn from any further analyses. The second reason was to take into account the small differences in TDS between counterbalances. Even though the distractors were held constant across each counterbalance, the target objects were different images which would impact the score. As expected, we found significant differences in the average TDS across counterbalances by cueing condition $F(2,4) = 62.423$, $p = .001$. Importantly, post-hoc tests revealed these differences followed the expected pattern where

subordinate trials had on average the highest TDS scores, followed by the basic, and then superordinate conditions (p's < .029). This finding indicates that TDS is a suitable replacement in the CCF model to the number of CCFs for a given category as subordinate trials seem to have more distinct targets.

For all further analyses, an average time-to-target and verification time was computed for each trial, drawn from correct responses from the participants. Conducting an overall correlation, collapsing across cueing condition, using every trials average time-to-target and TDS score revealed a significant correlation, $r(432) = -.164$, $p = .001$. Targets that were more distinct had lower time-to-target on average (Figure 13A). This result indicates that the CCF model is capable of capturing how fast attention was guided to the target object. However, when computing the correlations for individual cueing conditions, only the basic level correlation was significant $r(144) = -.257$, $p = .001$, while the subordinate $r(144) = -.13$, $p = .12$ and superordinate $r(144) = -.075$, $p = .372$, were not. This result shows that the strength of the basic level correlation was able to carry the overall correlation to be significant.

Turning to the models predictions of verification time, I found another significant overall correlation with TDS*Sibling Distance, $r(432) = .109$, $p = .024$, Figure 13B. In this case though, no cueing condition on its own reached a significant correlation (p's > .311). It would appear that while none of the conditions on their own had a large enough effect to reach significance, the consistent direction of the effect combined with an increase in observations was enough to reach significance in the overall prediction. Again, the CCF model is able to predict the average verification time on individual trials, albeit to a lesser extent. Identical analyses that were performed on individual categories instead of trials showed the same pattern of results.

Aside from the target present analyses, we were able to analyze the target absent results. Behaviorally, lures were more often fixated at the subordinate level both in terms of the proportion of fixations on the lure and dwell time. Because of this, another test of the CCF model is in predicting this additional data pattern. An overall correlation between TDS and the average time to fixate on the lure was significant, $r(216) = -.215$, $p = .001$. Additionally, the subordinate level correlation was significant $r(72) = -.384$, .001. Neither the basic or superordinate correlations were significant (p's > .714). This pattern of results demonstrates the CCF models ability to predict the preferential fixations on lure objects that are more similar to the target then either basic or superordinate lures.

The current work aims to extend the CCF model (Yu, Maxfield, & Zelinsky, 2016) to make predictions on the efficiency of search guidance when cued with a category. The core of the model remains the same. 100 images from each of the 48 subordinate categories were used as the training set. These same images were combined to create categories at the basic and superordinate level. SIFT (Lowe, 2004) and color features (van de Weijer & Schmid, 2006) were extracted for each image and then clustered into a 1,064 feature histogram using Bag-of-Words (BoW; Csurka, Dance, Fan, Williamowki, & Bray, 2004). Each bin in this histogram corresponded to a visual word, with the height representing the frequency of that visual word occurring in each individual image. These histograms were averaged together for each category. For each category's averaged histogram, bins that had too little frequency and/or too high variability were pruned away (Yu, Maxfield, & Zelinsky, 2016). The features that remained were called Category-consistent features, highly informative features that were present both frequently and consistently across the exemplars of its category. In previous work, we used the raw number of CCFs for each category to predict the trend in guidance times. For the present

experiment though, we are instead using the chi-squared distance of the CCF histogram to the BoW features for any given object in the search display. The lower this score is, the more similar the object's features are to the important features of the cued category.

I attempted multiple methods of analyzing the chi-squared distances for the objects in the search display. The resulting measure from these calculations I am calling the Target Distinctiveness Score. We can think of these methods as a miniature model comparison. Or more specifically, what algorithm would aid the CCF model in making predictions about guidance times for individual trials. Their descriptions and results are outlined below.

TargOnly – A correlation between the guidance measures and the chi-squared distance between the target category CCFs and the target in the search display. While this method yielded the strongest correlation, it also completely ignores the presence of other objects in the search display. While the distractors in this experiment were random objects with minimal feature overlap to the cued category, this measure would certainly fail to generalize in predicting performance on a search task with more target similar distractors.

Max – A correlation between the guidance measures and the difference between the target's chi-squared distance and the distractor with the lowest chi-squared distance. This method attempted to address the shortfalls of the target only approach. It is based on the idea that one of the primary determining factors for guidance is how similar any of the distractors are to the target. While the correlation with time-to-target was weaker than target only, it was still significant. The correlation being in the negative direction was expected since a positive difference score between the target's chi-squared distance and the most similar distractor indicated that the target stood out more and led to lower time to target on average. The opposite

is true for the correlation to first object fixation rates, a more positive value should lead to a higher proportion of trials in which the target was the first object fixated. This approach also suffers somewhat from the fact it only takes one of the other objects in the display into account. We could reasonably expect differences in guidance if there is one target-similar distractor compared to one hundred.

Average – A correlation between the guidance measures and the difference between the target's chi-squared distance and the average of all the distractor differences. Again we would expect a negative correlation value for time-to-target and a positive one for first object fixated. If the target is more similar to the cued category than the distractors, subtracting the target's distance from the average distance of all the distractors would result in a positive value. This positive value would be negatively correlated with time-to-target and positively with first object fixated. This measure yielded a somewhat higher correlation than Max. It also addresses the issue with the number of distracting objects by weighing each distractors distance equally. There was very little variance in the distances of the distractor objects, so it is unsurprising that the correlation was similar to the previous one. While this suites the current task where the eccentricity of the objects was held constant, it may perform worse in a task where eccentricity is not the same. The reason for this is that in addition to matching target features with objects in the display, another key factor in search guidance is retinal acuity limitations. Our visual system has a more difficult time recognizing objects in the periphery. This approach does not take this factor into consideration currently as it is only concerned with matching features between the objects in the display and the target category, regardless of eccentricity. This leads into the motivation for the next analysis which takes eccentricity into account.

Gaussian – A correlation between the guidance measures and the difference between the targets chi-squared distance and the average of all the distractor differences with a Gaussian noise distribution applied. The mean and standard deviation parameters of the Gaussian distribution were found by using a grid search to systematically test possible values. This noise could be attributed to retinal limitations for perceiving objects at greater eccentricity or neuronal noise. Adding it was a principled decision to drop the models results to closer approximate human performance levels. While participants first fixated on the target objects 27% of the time on subordinate trials (which had the strongest guidance), the CCF model in it's most basic instantiation predicted that the target should be fixated first 42% of the time (Figure 14). This analysis performed slightly better than just the previous averaging procedure, though it did not exceed just using only the target's distance. While less parsimonious, this measure would theoretically perform much better across a range of tasks that had objects at different eccentricities and with a broader range of target-distractor similarity. This is because the Gaussian distribution would simulate retinal acuity limitations while the comparison of each object to the cued CCF histogram would approximate feature similarity.

Regression – A linear regression was performed attempting to predict the guidance measures using the distances from each of the six objects as predictors. The result for time-to-target was significant, $F(1,430) = 5.13$, $p = .024$, with only the distance of the target object having a significant coefficient to effect the finding ($t = 2.27$, p = .024). This makes sense for the same reasons as stated earlier, there was not much variance in the distractor objects and the target alone correlation was the strongest. One main advantage to this analysis is that it does not require the supervised approach of having the experimenter select out what the target object was for the TDS to be computed. This analysis simply takes all of the distances for a given trial and

attempts to predict the guidance measures (which are still defined by knowledge of which object is the target).

All the analyses performed have benefitted from using measures where the target object was known (time to target and first object fixation rates are computed based off this information). An alternative approach could be to forego this information in favor of a more bottom-up methodology where comparisons are made based on where participants tended to look as opposed to if an object was a target or not. For example, it could answer the question of whether the chi-squared distance of the objects that were fixated could predict guidance measures. Since the goal of the current work is to predict gaze behavior, it makes sense to use that same behavior to define the groups for comparison. This would be a stricter test of the model's ability to predict guidance behavior, but one that the current CCF model would fail. While the CCF model may not currently be able to account for behavior at this level, it remains a possible goal for models of visual search.

It should be noted that none of the analyses were significant in regards to the proportion of trials in which the target was the first object fixated. Capturing this averaged proportion is a difficult task since categorical search is known to be weaker than pictorial. The average for this measure was only ~5% higher than chance levels. This compressed value range could be the reason we are seeing such poor results when examining this specific measure of guidance.

Discussion

Object categories likely differ with respect to the "goodness" of their visual representations in guiding behavior. Recently, we introduced a CCF model that predicted search guidance to targets at different levels of the category hierarchy. Through a series of correlations,

it has been shown that the CCF model is able to predict the average time-to-target for target-present basic trials, verification time when all trials are taken into account, as well as the capture of attention by subordinate level lures on target-absent trials. This work extends our CCF model to the prediction of guidance and verification on individual trials.

The correlation for guidance at the basic level was a somewhat surprising result. Even though subordinate cues had lower time-to-target on average, the analyses revealed that the basic level was more predictable. This cannot be accounted for by differing levels of variance in guidance times as the variance in the behavioral data was nearly equal (standard deviations of 201 and 212 ms respectively). It is possible then that two aspects of the TDS would lend favor to the basic level, either in the calculation of the chi-squared distance to the cues CCF histogram or in the comparison of the distractors distances. An analysis using just the chi-squared distance of the target to the cue revealed a similar correlation. This suggests that the CCF model may just be doing a better job of matching what the important features are at the basic level. Future work will aim to extend and improve upon the CCF feature selection through deep learning methods with parameters informed by the ventral stream.

In terms of the verification correlation, the consistent trend of the smaller effects at each level combined to make a small but significant correlation present when all trials were compared together. This pattern of results is likely a result of the smaller effect size in the behavioral data. In all of our previous studies, the guidance effect had was usually at least twice the magnitude of the verification differences. While we were able to reliably differentiate these conditions using precise examination of eye movement data, the correlation was not able to account for the smaller differences that separated each cueing condition. The CCF model performed as expected though when predicting the target-absent trials, where the effect size was again substantial.

These results in particular speak to the importance of including distractors, lures and non-lures alike, into any model of search performance.

It would be possible in future work to use the categorical target map generated by the CCF distances as the input for a model such as MASC. However, directly comparing performance between the two would be a sizeable undertaking. Whereas the CCF model dimensionally reduces the feature histograms of categories, MASC and most other models that leverage the computation of visual features use the whole unaltered histogram of a learned category. These histograms are usually incredibly large and filled with more features than we could possibly load into the target template we use to guide search. By using every bit of information in order to aid in differentiating categories, these dense representations can then be weighted or passed through noise in order to fit the behavioral data. The CCF model on the other hand is making a stronger claim and attempting to specifically identify which features are important for a given category and using this information to predict human search performance. So while I would expect MASC to potentially have a better fit to any behavioral data, it should not undercut the contributions provided by the CCF model.

General Discussion

The work done for this dissertation was focused on exploring categorical hierarchies from multiple angles. Experiment one manipulated the set size of a categorical search task to test the notion that cues of varying hierarchical levels are differentially effected by additional distractors in the display. Overall, the experiment provided evidence that the basic-level advantage previously observed in a categorical search task may be more fragile than anticipated, being very susceptible to task demands and inter-object distances. Experiment two attempted to identify

specific neural ERP components to differentiate hierarchical categorical cues during a search task. Unfortunately, no significant differences emerged between cueing conditions, making conclusive determinations difficult. Experiment three analyzed the strengths and weaknesses of two different methods for calculating similarity between complex objects, MDS & CCF model. While neither were able to correlate with the reaction time of a category verification task, there is much to be learned from contrasting these two methods. Namely, that both are capable of determining centroids of categories that could be useful in predicting performance on other types of tasks. Experiment four sought to refine the CCF models methods to work on the finer scale of predicting individual guidance and verification times during a search task. A target map was generated using the distance from each object in the search display which was then used to successfully account for a segment of the data.

As mentioned previously, there are a number of hurdles to overcome for research focused on categories. The first and arguably most important is the enrichment of the stimuli to include photorealistic objects. While the bulk of research on learning categories has necessarily used impoverished stimuli for controls sake, we have demonstrated the capability of new methods that can handle complex objects. The doors are now open for researchers to use databases consisting of thousands of categories with exponentially more exemplars. The knowledge gleaned over previous decades is due to be tested against ecologically valid methods. This extends also to examining the neural mechanisms that underlie common tasks such as categorical search.

Categories are complex structures, but we have new tools to explain old ideas. It is possible to scale up our understanding of how categories are represented and used in order to take a step closer to predicting behavior in the real world. By leveraging computational models

that are by their definition functionally explicit, there is a new platform from which to explore and make predictions about behavior.

References

Adeli, H., Vitu, F., & Zelinsky, G. (2015). A model of saccade programming during scene viewing based on population averaging in the superior colliculus. Abstract presented at Vision Sciences Society. *Journal of vision*, *15*(12), 365-365.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22, 261–295.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 38, 423–466.

Batty, M., & Taylor, M. J. (2002). Visual categorization during childhood: an ERP study. *Psychophysiology*, *39*(4), 482-490.

Bundesen, C. (1990). A theory of visual attention. *Psychological review*, *97*(4), 523.

Chelazzi, L., Miller, E. K., & Duncanf, J. (1993). inferior temporal cortex. *Nature*, *363*, 27.

Cohen, M. R., & Maunsell, J. H. (2011). Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron*, *70*(6), 1192-1204.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological review*, *96*(3), 433.

Drew, T., Vo, M. L-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again, sustained inattentional blindness in expert observers. *Psychological Science, 24(9),* 1848-1853.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and clinical neurophysiology*, *99*(3), 225-234.

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in cognitive sciences*, *18*(10), 526-535.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112.

Hagen, S. Vuong, Q., Scott, L.S., Curran, T., & Tanaka, J.W. (2014) The role of color in expert object recognition. *Journal of Vision, 14(*9):9, 1-13.

Hoffman, J., Guadarrama, S., Tzeng, E. S., Hu, R., Donahue, J., Girshick, R., Darrell, T., & Saenko, K. (2014). LSDA: Large scale detection through adaptation. In *Advances in Neural Information Processing Systems* (pp. 3536-3544).

Hopf, J. M., Luck, S. J., Girelli, M., Hagner, T., Mangun, G. R., Scheich, H., & Heinze, H. J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, *10*(12), 1233-1241.

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2012). Multidimensional scaling. *Wiley Interdisciplinary Reviews (WIREs): Cognitive Science, 4*, 93-103.

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256.

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology*, *5*(1), 1-10.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 4-4.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 829– 846.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling* (Vol. 11). Sage.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review*, *9*(1), 43-58.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision research*, *48*(5), 635-654.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. Psychological Review, 111, 309–332. Lowe, D. G. (2004). Distinctive image

features from scaleinvariant keypoints. *International Journal of Computer Vision, 60*, 91–110.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 1000.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics to neurobiology. *Trends Cogn. Sci.* 17, 391-400.

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632-1639.

Mack ML and Palmeri TJ (2011) The timing of visual object categorization. *Front. Psychology 2*:165

Mack, M. L., Wong, A. C. N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2009). Time course of visual object categorization: Fastest does not necessarily mean first. *Vision research*, *49*(15), 1961-1968.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision,* 9 (11): 8, 1–13,

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*(9), 744-751.

Maxfield, J. T., Stadler, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision, 14(*12), Article 1. doi:10.1167/14.12.1

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition, 20*, 1153–1163.

Mecklinger, A., & Ullsperger, P. (1993). P3 varies with stimulus categorization rather than probability. *Electroencephalography and clinical neurophysiology*, *86*(6), 395-407.

Meuter, R. F. I., & Lacherez, P. F. (2015). When and why threats go undetected: Impacts of event rate and shift length on threat detection accuracy during airport baggage screening. *Human Factors: The journal of human factors and ergonomics.*

Murphy, G. L. (2002). The big book of concepts. Cambridge, MA: MIT Press.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 70–84.

Nako, R., Wu, R., & Eimer, M. (2014). Rapid guidance of visual search by object categories. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 50.

Nako, R., Smith, T. J., & Eimer, M. (2015). Activation of new attentional templates for real-world objects in visual search. *Journal of cognitive neuroscience*.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387-391.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology, 75*, 97-129.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological review*, *104*(2), 266.

Pazzani, M. (1991). The influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 416–432.

Polich, J., & Comerchero, M. D. (2003). P3a from visual stimuli: typicality, task, and topography. *Brain topography*, *15*(3), 141-152.

Rosch, E. H. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology, 62*, 1904–1914.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Normed for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Tanaka, J., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event□related potentials. *NeuroReport*, *10*(4), 829-835.

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.

van de Weijer, J., & Schmid, C. (2006). Coloring local feature extraction. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II (Lecture Notes in Computer Science, Vol. 3952, pp. 334–348). Berlin, Germany: Springer.

Vickery, T. J., King, L., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5,* 81-92.

Wascher, E., Hoffmann, S., Sänger, J., & Grosjean, M. (2009). Visuo□spatial processing and the N1 component of the ERP. *Psychophysiology*, *46*(6), 1270-1277.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160-168.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44*, 1411-1426.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*(6747), 867-869.

Wu, R., Scerif, G., Aslin, R. N., Smith, T. J., Nako, R., & Eimer, M. (2013). Searching for something familiar or novel: Top–down attentional selection of specific items or object categories. *Journal of cognitive neuroscience*, *25*(5), 719-729.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49¸2095-2103.*

Yu, C-P., Maxfield, J.T., & Zelinsky, G.J. (2016). Searching for category-consistent features: A computational approach to understanding visual category representation. *Psychological Science, 27*(6), 870-884.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological review*, *115*(4), 787.

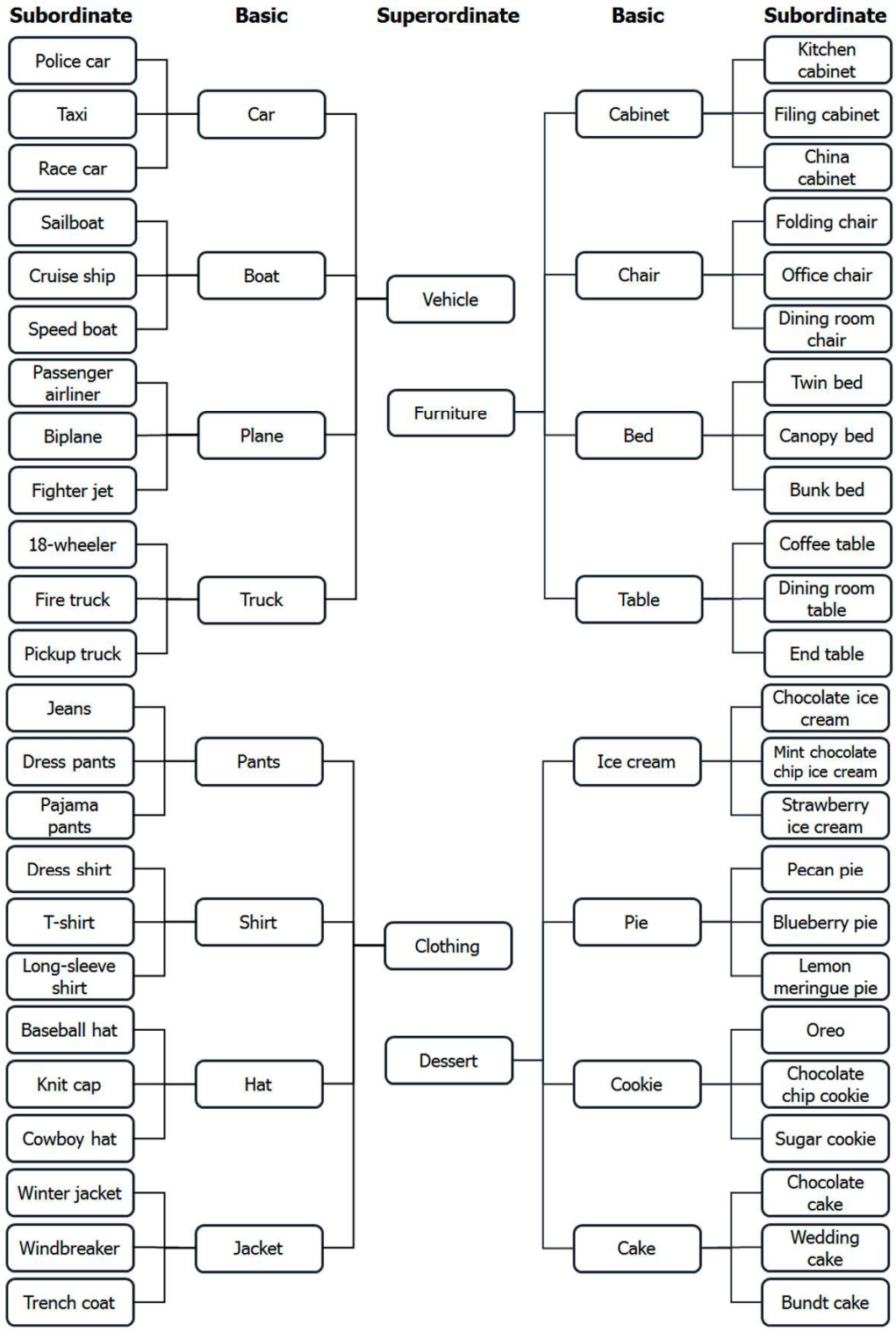
Zelinsky, G. J., Adeli, H., Peng, Y., & Samaras, D. (2013). Modelling eye movements in a categorical search task. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*, 20130058.

Zweig, A., & Weinshall, D. (2007, October). Exploiting object hierarchy: Combining models from different category levels. In *2007 IEEE 11th International Conference on Computer Vision* (pp. 1-8). IEEE.

Figures

Figure 1. (A) The hierarchy of categories used in Experiments 1 – 4. (B) A sample of targets and distractors used in Experiments 1-4.

A)



B)

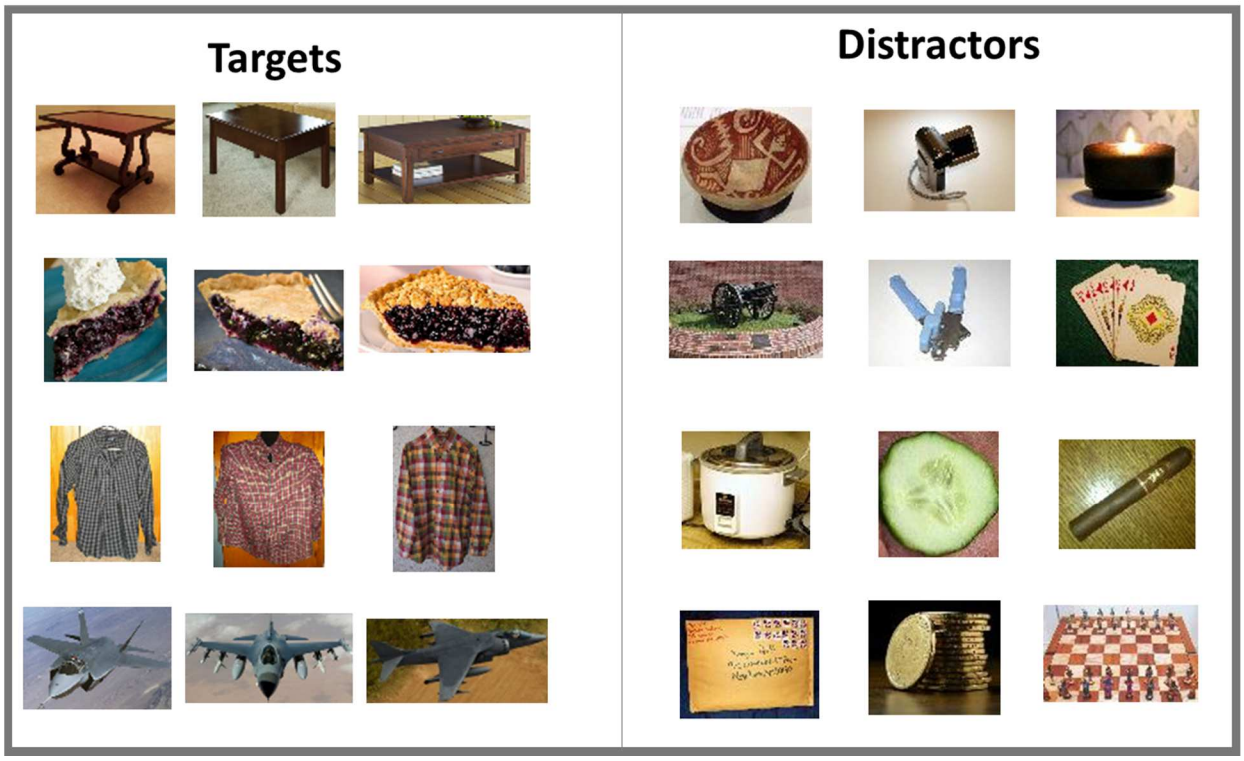


Figure 2. Procedure from Experiment 1.

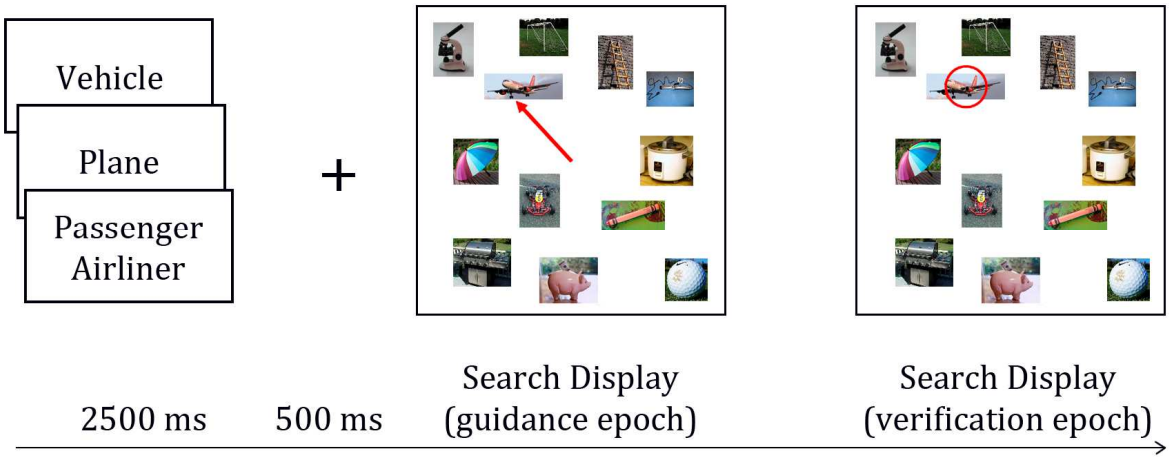
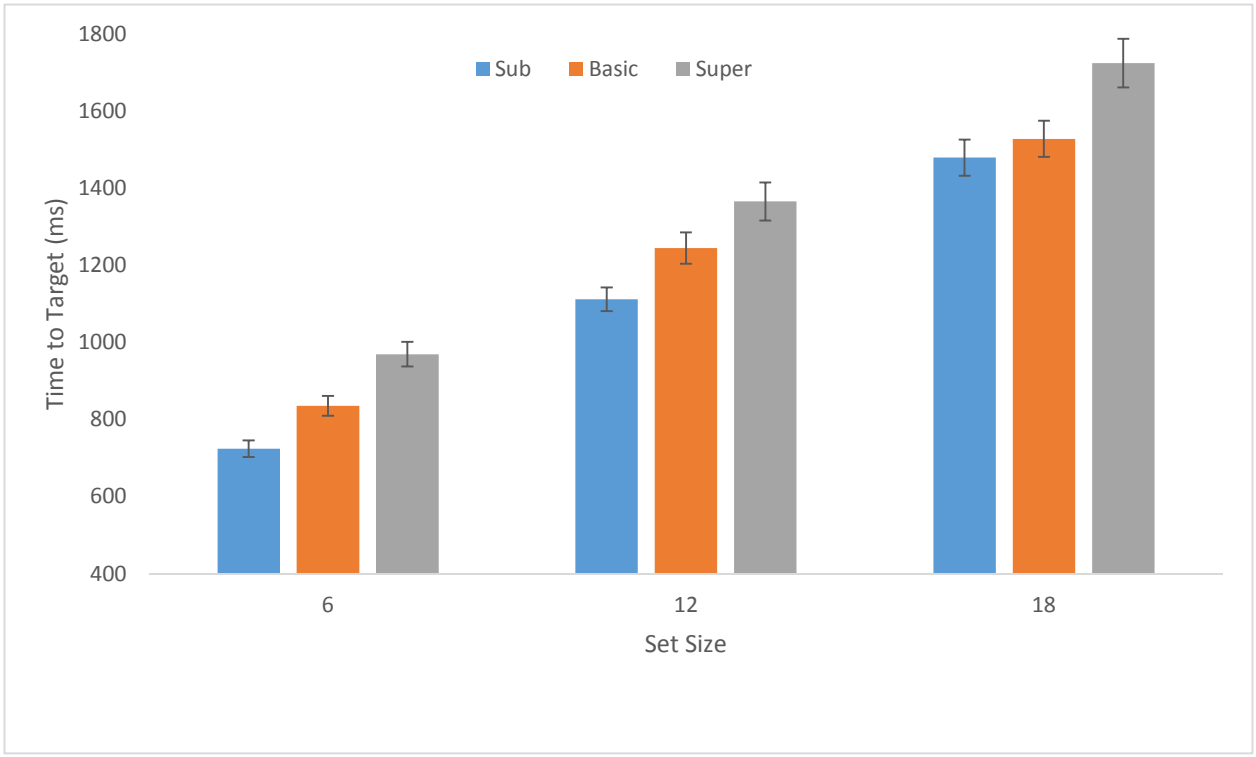


Figure 3. (A) Time-to-Target and (B) Verification times for correct target-present trials broken down by set size and hierarchical level in Experiment 1.

A)



B)

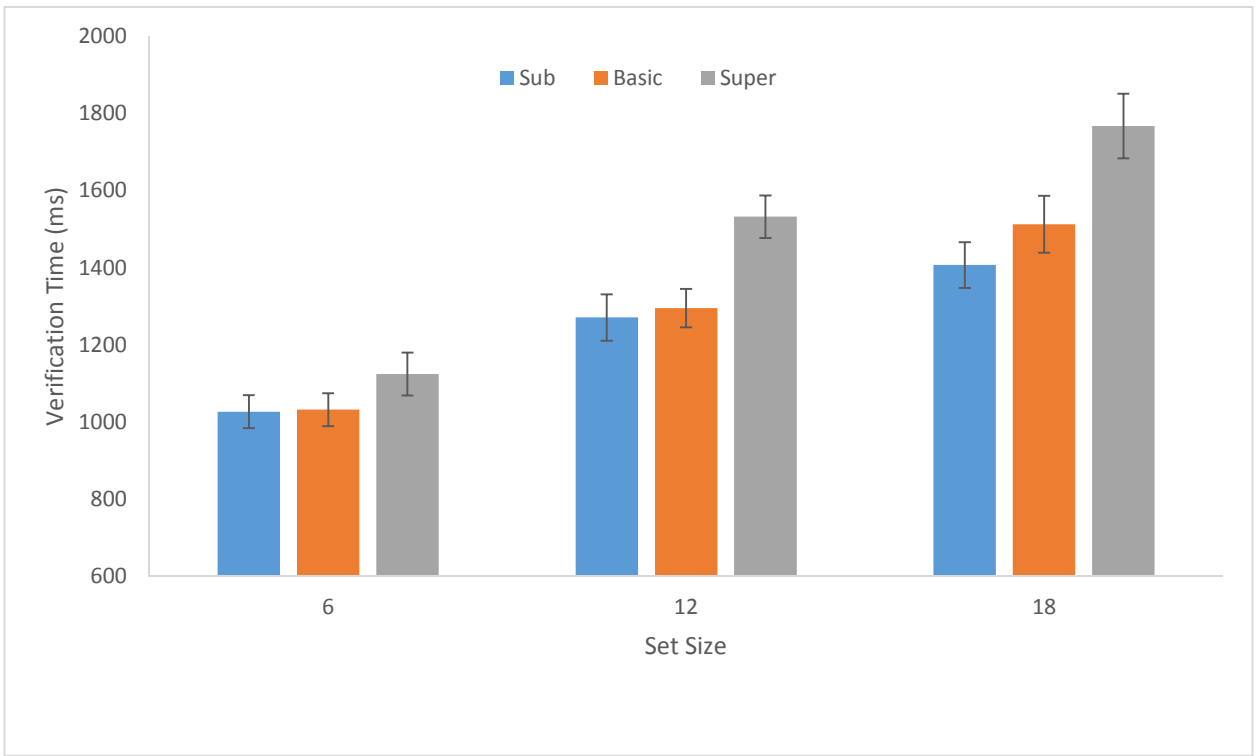


Figure 4. Procedure from Experiment 2.

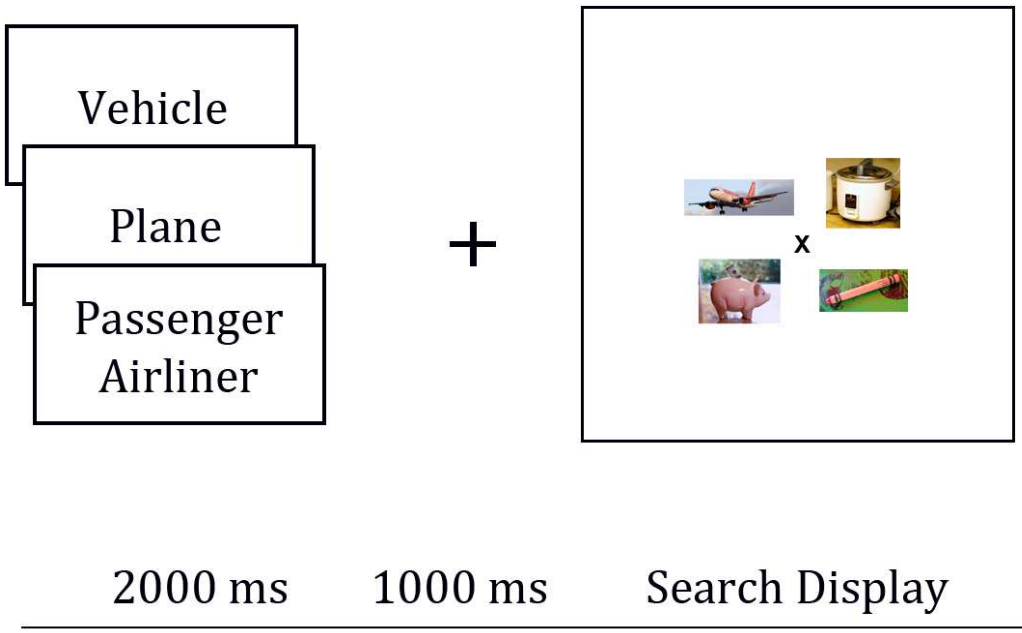


Figure 5. Reaction times from correct target-present trials in Experiment 2.

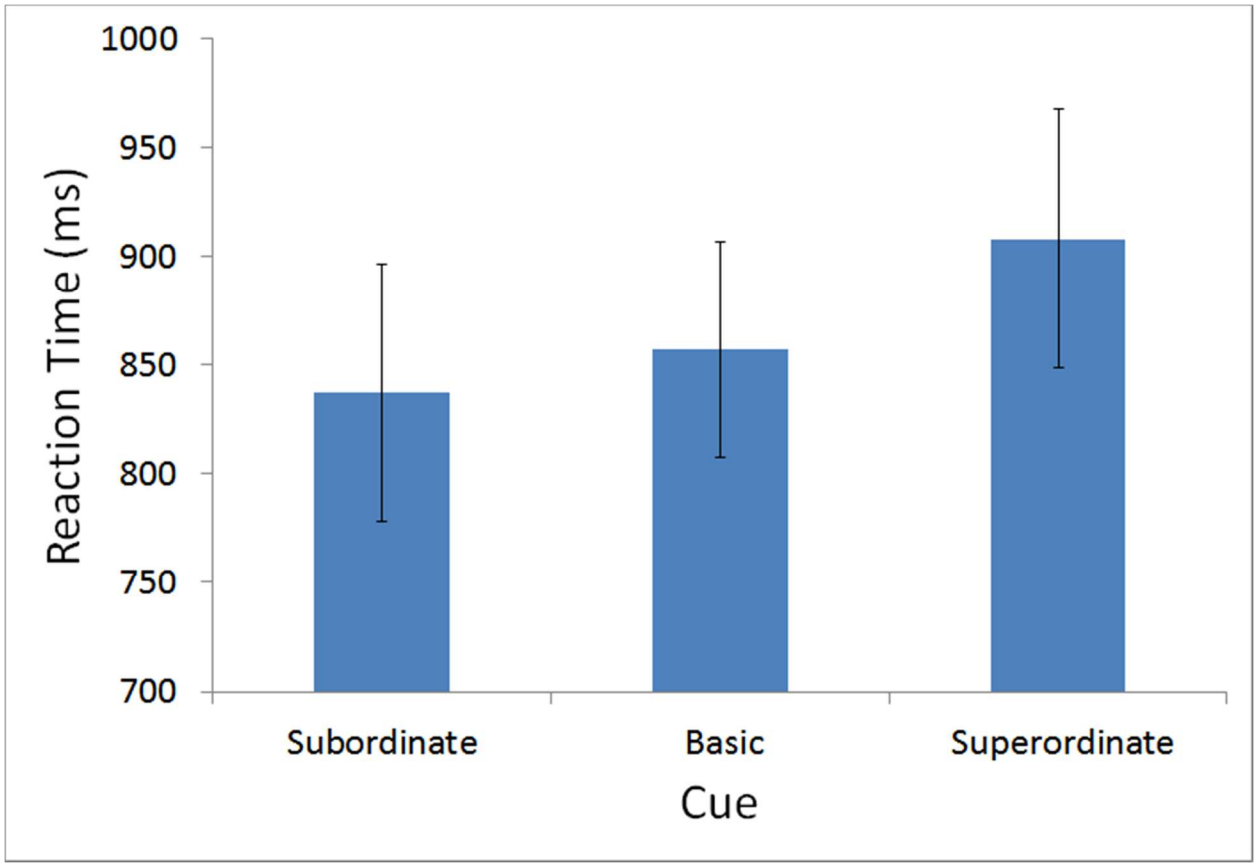


Figure 6. Headmaps showing contralateral minus ipsilateral EEG activity for time window between 150 – 500 ms post stimulus onset for (A) subordinate, (B) basic, and (C) superordinate cues.

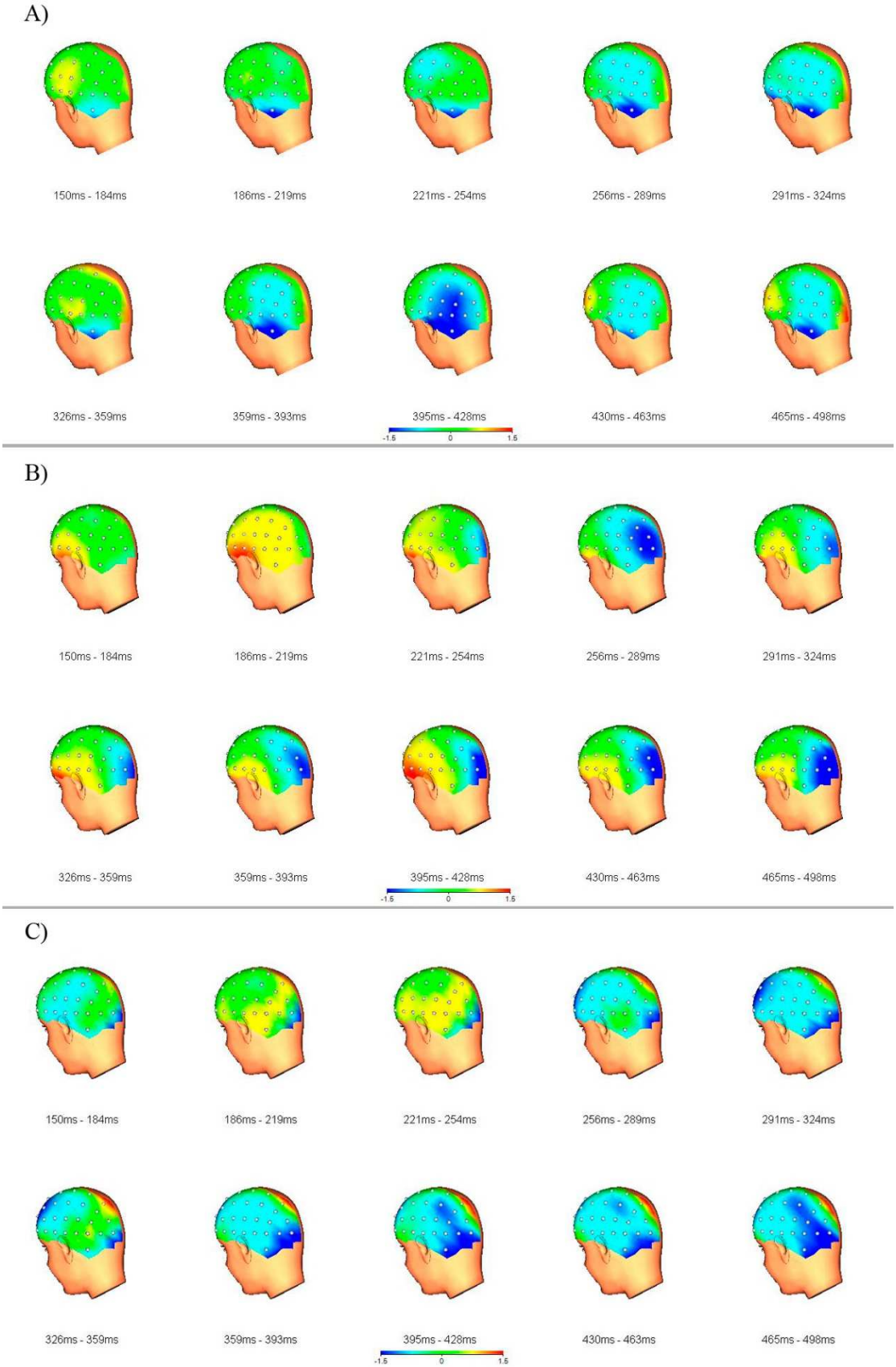


Figure 7. An example of a SpAM trial used in Experiment 3.

Figure 8. An overview of the Category-Consistent Features Model. Taken from Yu, Maxfield, & Zelinsky, 2016. (i) 100 images of object exemplars were collected for 48 subordinate-level categories. These exemplars were combined to create 16 basic-level categories (each with 300 exemplars) and 4 superordinate-level categories (each with 1200 exemplars). (ii) SIFT and color histogram features were extracted from each exemplar and the Bag-of-Words (BoW) method was used to create from these a common feature space consisting of 1064 "visual words". (iii) 1064-bin BoW histograms were obtained for each exemplar, where the bins correspond to the visual words and bin height indicates the frequency of each feature in the exemplar image. BoW histograms were averaged by category to obtain 68 averaged histograms, each now having a mean frequency and variability associated with each visual word. (iv) Features in these averaged histograms having too low of a frequency or too high of a variability were excluded, resulting in a lower-dimensional feature representation of each category that we refer to as category-consistent features (CCFs)—those highly informative features that are present both frequently and consistently across the exemplars of a category.

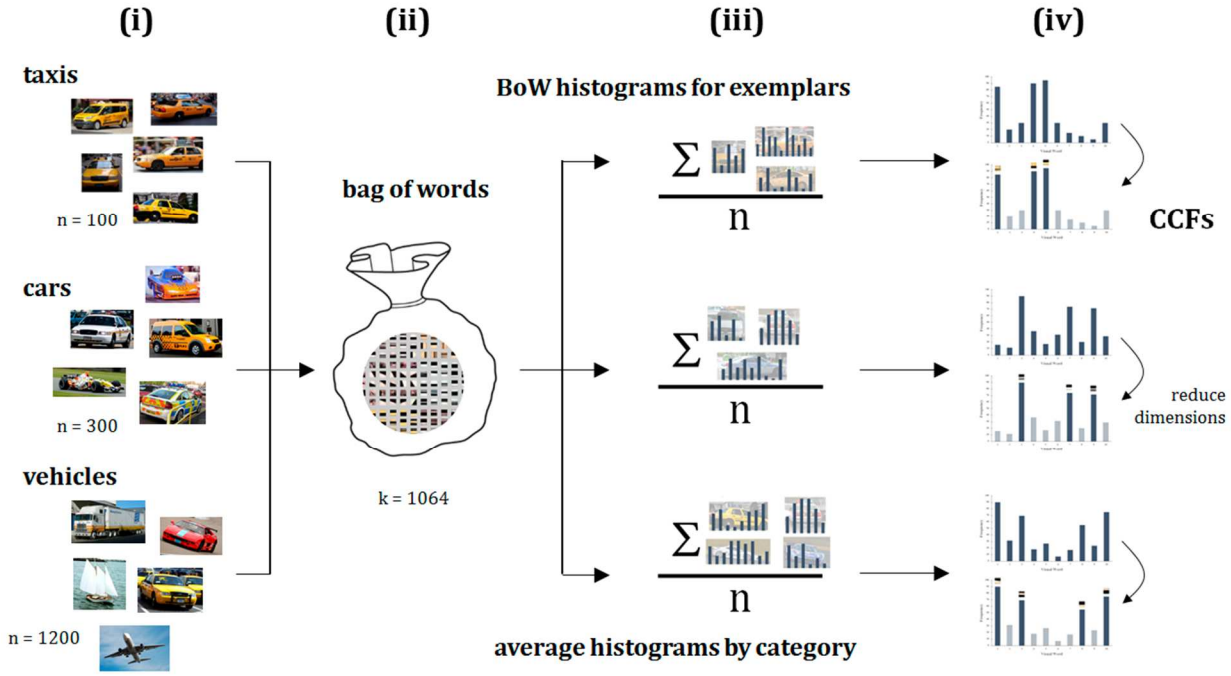


Figure 9. The average inter-item distance for each hierarchical level derived from MDS.

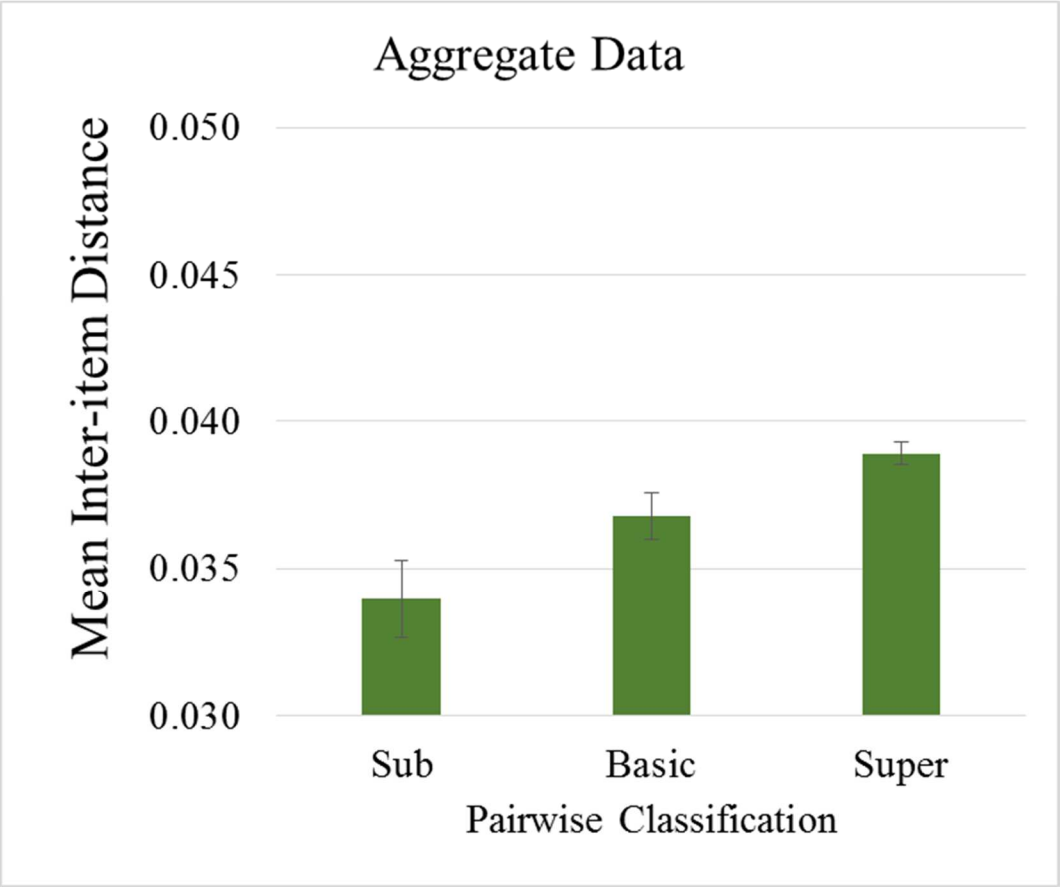


Figure 10. Reaction times for correct target-present trials for the category verification task in Experiment 3.

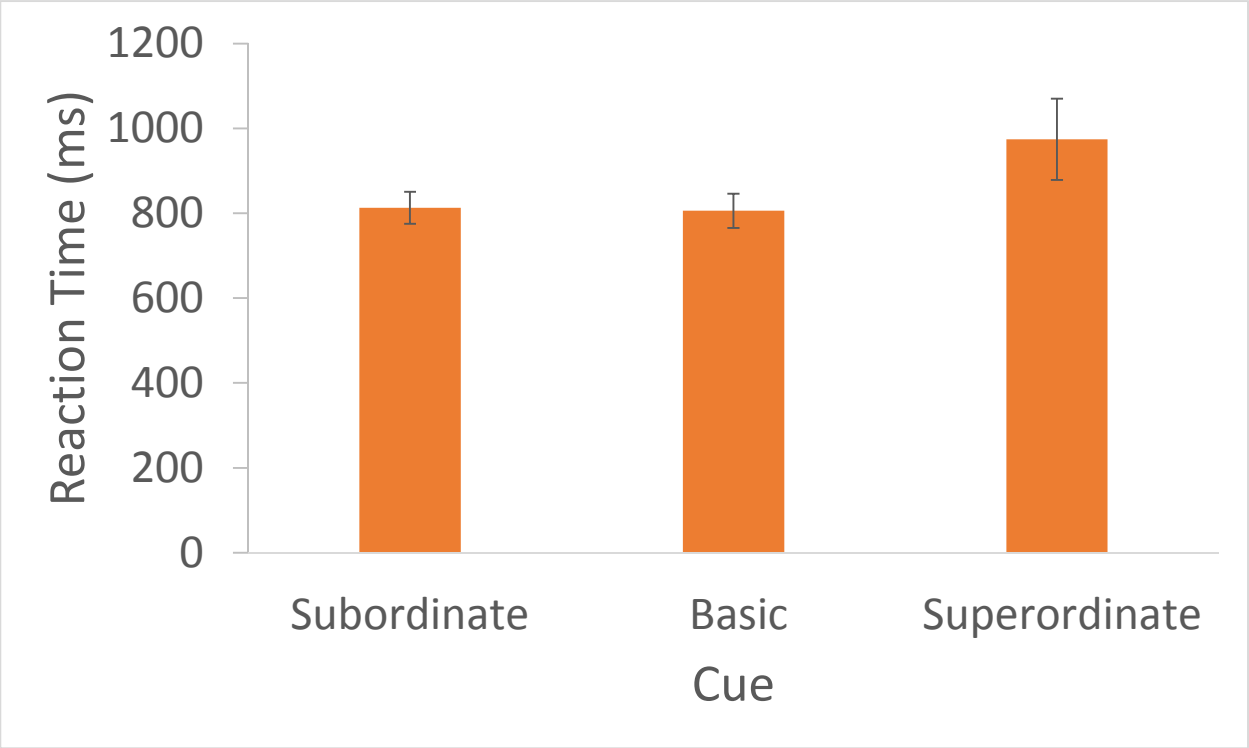


Figure 11. An illustration of how the chi-squared distances between the CCF histogram of the cue and each object in the search display were generated in Experiment 4.

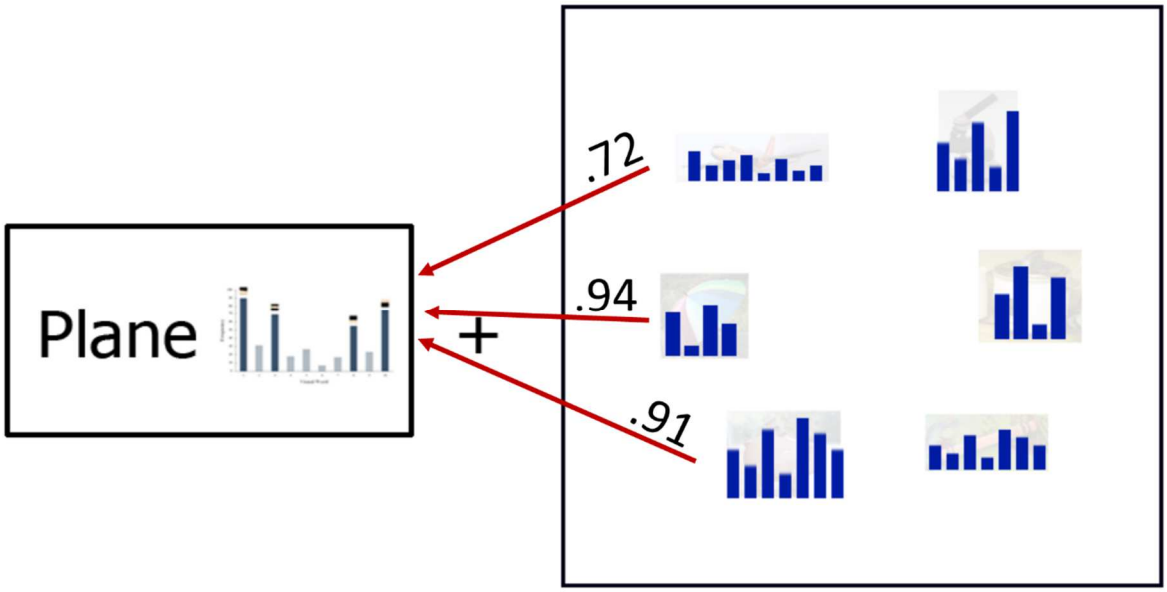
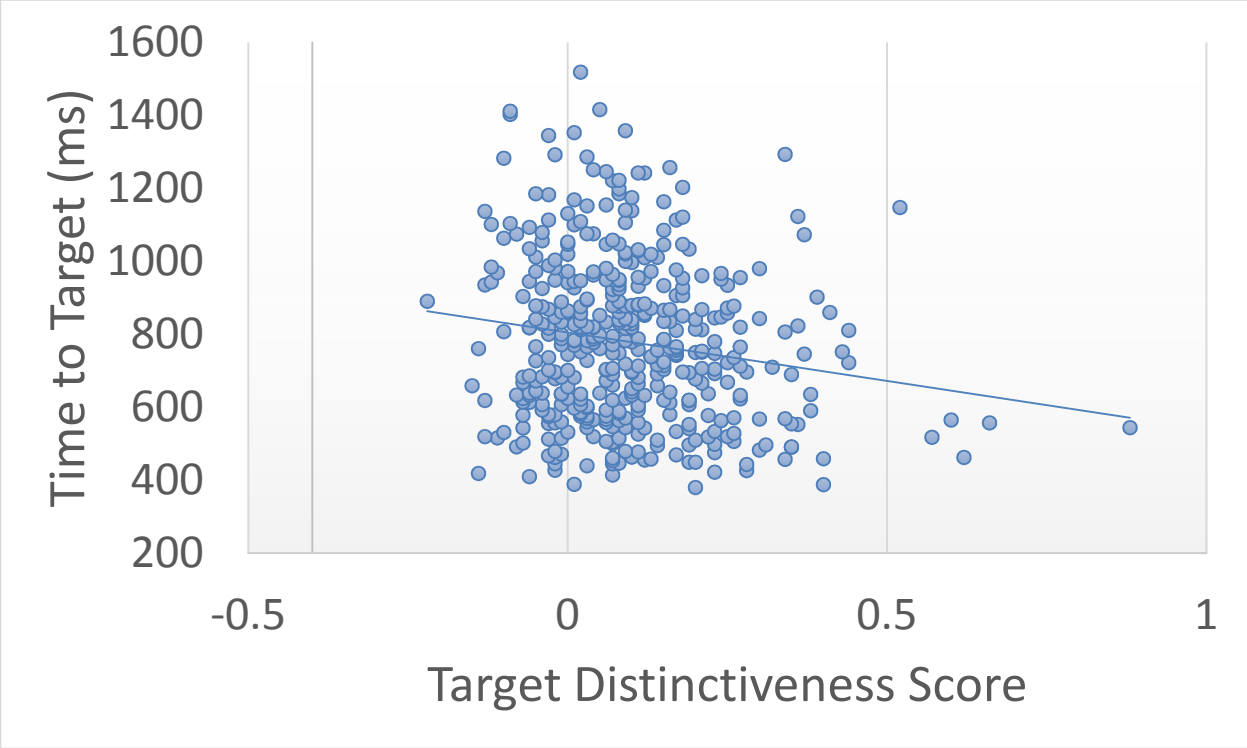


Figure 12. An example of the target map generated in Experiment 4 where intensity

Figure 13. Correlations between the CCF model predictions on individual trials between (A) time-to-target and target distinctiveness score and (B) verification time and (target distinctiveness score * sibling distance).

A)



B)

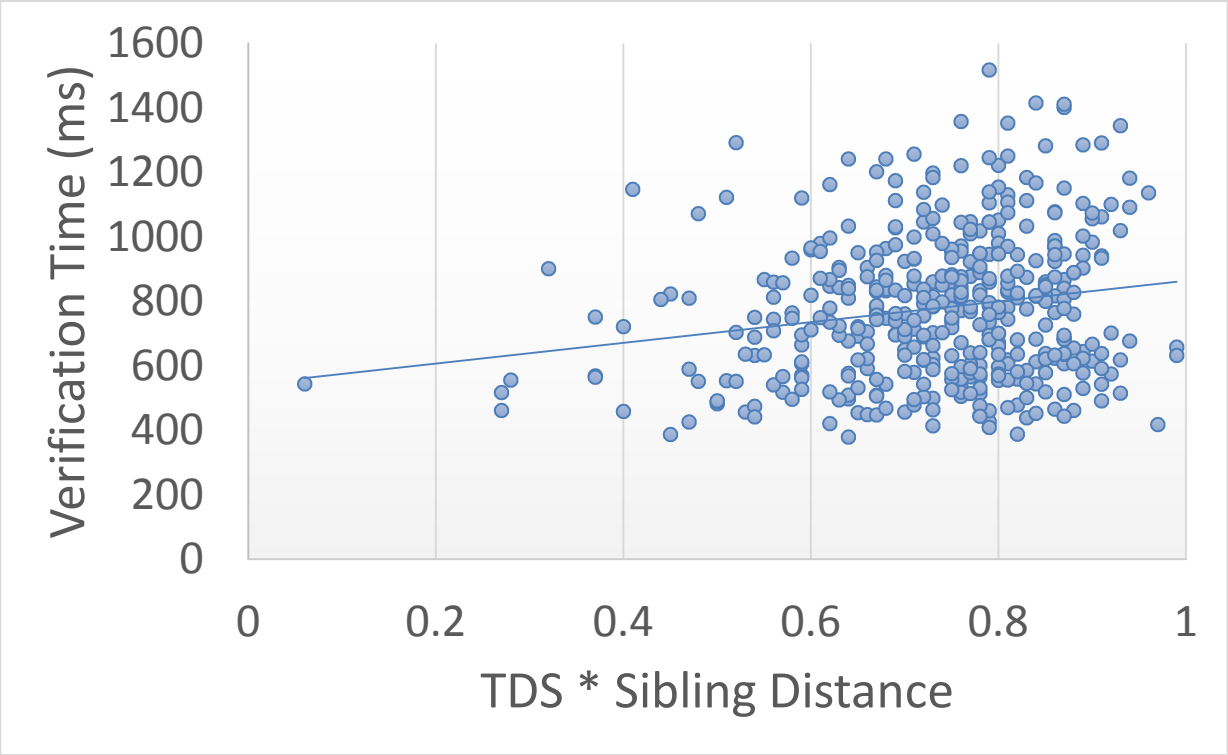
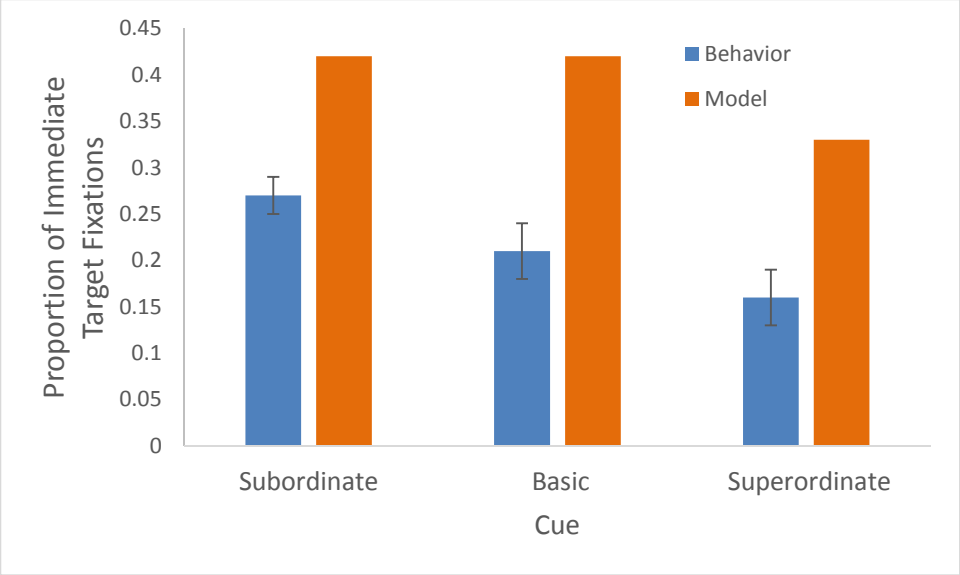

Verification Time (ms)
1600
1400
1200
1000
800
600
400
200
0
0
0.2
0.4
0.6
0.8
1
TDS * Sibling Distance

Figure 14.

A comparison of the behavioral guidance data to base CCF model predictions without any additional algorithms. The model predictions are generated by predicting that the target should be the first object fixated if it has the lowest chi-squared distance to the cued category's CCF histogram. The fit for this rudimentary model is poor, with the model overestimating human levels of performance. This figure demonstrates the need for an additional step to refine the models predictions.



Tables

Table 1. A summary of the correlations comparing the various measures attempted to have the CCF model make predictions on individual trials.

	Time-to-Target		**First Object Fixated**	
Model Type	Statistic	p-value	Statistic	p-value
CCF + TargOnly	r = .176	.001	r = -.03	.541
CCF + Max	r = -.114	.018	r = .043	.375
CCF + Average	r = -.164	.001	r = .031	.523
CCF + Gaussian	r = -.17	.002	r = .029	.544