2018

Visual Search in Naturalistic Imagery

Dave Schreifels
Michigan Technological University, djschrei@mtu.edu

VISUAL SEARCH IN NATURALISTIC IMAGERY

By

Dave J. Schreifels

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

This thesis has been approved in partial fulfillment of the requirements of the Degree of MASTER OF SCIENCE in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Thesis Advisor: *Shane T. Mueller*

Committee Member: *Kelly S. Steelman*

Committee Member: *Robert L. Pastel*

Department Chair: *Susan L. Amato-Henderson*

Table of Contents

Index of Figures

Acknowledgments

This thesis document is the culmination of 1.5 years of research and work at Michigan Technological University. The road to this point has been long and winding, and I wouldn't have been able to make it this far without the influence of several key groups of people in my life. I think it appropriate to address each of these groups in turn.

I extend my deepest gratitude to the people who supported me during what was, for me, the hardest part of graduate school – the beginning. This was a very dark time for me, and the following people were instrumental in my perseverance during this time:

Anastasia Rogers and Travis Lockhart

as well as Andrew Summers; Ryan Bancroft; Sterling Korstad; Donald Williams; Michael Irish; Michael Christianson; and Shane Mueller

To all of you listed above: Thank you, from the bottom of my heart. In one way or another, each one of you was willing to help in a major way at one or several critical points (which were nothing short of frequent) during the most challenging time of my life. There are no words to express how grateful I am for the positive impact you've had on my life.

I extend my gratitude to my parents, Kathleen and Joseph Schreifels, as well as my girlfriend, Brittany Nelson, who have been extremely loving and supportive even when I have made myself difficult to love and support. It is only thanks to you that I have grown into the man I am today. Thank you for helping me to be a more patient and caring person in my everyday life.

I additionally extend my gratitude to my advisory committee – consisting of my advisor, Shane Mueller; Kelly Steelman; and Robert Pastel – for their patience and feedback over the last several months. *To address Shane in particular:* Your approach to advising has made great hurdles seem like minor bumps on the long and winding road to making my first major contribution to the scientific literature. Thank you for all of the little things you do behind the scenes that have made graduate school so much more bearable. It has been an honor.

Finally, I would like to thank my other friends, professors, and colleagues of the last six years. While there are too many of you to name, your camaraderie, advice, and good humor have proved indispensable throughout my time at Michigan Tech.

This was only possible because of all of you. Thank you.

List of Abbreviations

AET	Attentional Engagement Theory
BC	Background Complexity
FIT	Feature Integration Theory
GS	Guided Search
NN(S)	Nontarget-Nontarget (Similarity)
PEBL	Psychology Experiment Building Language
RT	Response Time
TB(S)	Target-Background (Similarity)
TD(S)	Target-Distractor (Similarity)
TN(S)	Target-Nontarget (Similarity)

Abstract

Visual search has been extensively studied in the laboratory, yielding broad insights into how we search through and attend to the world around us. In order to know if these insights are valid, however, this research must not be confined to the sanitized imagery typically found within the lab. Comparatively little research has been conducted on visual search within naturalistic settings, and this gap must therefore be bridged in order to further our understanding of visual search. Based on the results of Experiment 1, Experiment 2 was conducted to test three common effects observed in previous studies of visual search: the effects of background complexity, target-background similarity, and target-distractor similarity on response time. Results show that these hypotheses carry over to the natural world, but also that there are other effects present not accounted for by current theories of visual search. The argument is made for the modification of these theories to incorporate this naturalistic information.

1 Introduction

Visual search is something we do every day in every part of our lives. Whether it's looking through the pantry for something to eat or watching out for drivers, road signs, and pedestrians on our morning commute, we search for objects and people that are relevant to our current activities. This doesn't just happen by magic, though; something is responsible for directing our attention to interesting and promising locations – for deciding when to keep searching for a target and, sometimes more importantly, when to stop.

This kind of research has broad applications; if a mechanism permeates our everyday lives, it stands to reason it will influence much of our lives. For example, as opposed to a century ago, it is now common to own an automobile and drive it nearly everywhere you go, even if it's just down the street. This is done at speeds the human body is not typically capable of withstanding, and so safety features have been implemented, including seatbelts, airbags, and, perhaps less commonly thought of, road signs. Indeed, road signs are crucial to safe driving; they moderate when to go and to stop, when to speed up because there's nothing but trees until the next city or to slow down because you've just entered a neighborhood where children run into the street without warning. But if we couldn't pick these road signs out of the background scenery, it would be little better than having no road signs at all (Ruff, 1950).

The concept of, for example, a stop sign sticking out of the background based on its features (i.e., a red sign against a green background) is commonly called "saliency." Saliency can be impacted by such things as the object's features and the context in which it appears. A stop sign is designed to be more "salient" by using bright colors not often found in nature, which makes you more prone to notice the sign where it is posted (Mogelmose, Trivedi, & Moeslund, 2012). Consider if stop signs were green in rural areas, or brown or grey in urban environments; in these situations the stop sign might be less salient due to its context.

While saliency should not be conflated with search, the two do go hand in hand. An object that is not salient will be harder to search for; similarly, one can look directly at an object one is searching for without seeing it (Mack & Rock, 1998). Noticing the stop sign sticking out of unfamiliar scenery is not visual search, but rather results from saliency; however, if one is searching for the stop sign because one was told it is there, perhaps by another road sign, saliency will impact how quickly and easily the sign is found. Saliency describes how the size, placement, and color of stop signs make them stick out, and it's why they have the same design almost anywhere you go. Toward this end, the principles learned from visual search research (sometimes referred to more broadly as "vision research") can guide our understanding of what will make the sign more salient, and what will have the opposite effect.

It is perhaps obvious that this kind of research is not limited to driving safely. For example, forensic investigators are always visually searching through crime scenes for evidence, and the saliency of that evidence can make the difference between finding it or not. The black light is a good example of this; with this technology, investigators are able to efficiently look for several types of organic evidence without having to closely inspect every nook and cranny of a building or room, because the black light makes certain organic compounds fluoresce and stand out from the background (Virkler & Lednev, 2009). This speeds the search for these kinds of evidence, which increases the efficiency of the investigation and, by extension allows for more investigations to be carried out.

There are applications in radiology as well. Radiologists must search for cancer spots within hundreds or thousands of clean scans, and the spots aren't always easy to make out. In breast cancer screenings (a semi-annual examination), those who are diagnosed with cancer have until that point been given a clean bill of health. However, for a significant number, the "clean" scans from previous years will show the cancer spots, indicating a problem of missed targets, or false negative errors (Bird, Wallace, & Yankaskas, 1992; Kundel & La Follette Jr, 1972). The Transport Security Administration faces a similar issue. They search for guns, bombs, and other threats in thousands of clean suitcases every day, but they're prone to the same problems as radiologists detecting cancer – there is high potential for false negative errors, and these can obviously be very dangerous (Wolfe, Brunelli, & Rubinstein, 2013). There exist years of research on these very issues, and each breakthrough could make the difference for thousands of people.

These examples have largely focused on the "bottom-up" aspect of saliency, or what features of the target and its context make it more or less salient. It should be noted, however, that these tasks are not entirely bottom-up processes; expertise helps guide TSA inspectors and radiologists in ways that novices are not equipped to do. This is an example of “top-down” influence, or the tendency of saliency to be influenced by the expectations of the observer. Let us again examine the stop sign example from a top-down perspective. Suppose a residential intersection has a four-way stop on it, but one of the signs has become partially obscured by leaves over time, thereby becoming less salient. Residents of the neighborhood will know the stop sign is supposed to be there, and their expectation will make the now-obscured sign easier to find through the foliage simply by virtue of their expectation that it is there (if it is even found at all – they may simply stop out of habit; Oliva, Torralba, Castelhano, & Henderson, 2003). However, a driver new to the area – a relative from out of town, perhaps – would not know about the obscured stop sign and may fail to find it before driving through the intersection. A driver who does not stop when other drivers expect them to stop can of course be extremely hazardous (Most & Astur, 2007), and this is just one example of the top-down influence on saliency. While this paper does not dispute the effects top-down processing on target saliency, these effects are somewhat beyond the scope of this research.

These applications of visual search are but a small sampling of some of the real-world contributions brought forth by research on visual search. Largely, these contributions are based on testing changes in certain observable effects which recur within the literature. Three robust, recurring effects are the focus of this research: target-distractor similarity (i.e., how similar a target is to other “candidate” targets), target-background similarity (i.e., how similar a target is to the background which camouflages it), and background complexity, which I operationalize as how much clutter is in the background. These three factors have been repeatedly shown to impact response time on several different visual search tasks, and they are derived from prominent theories of visual search which will be discussed in the next chapter. Notably, all three of these effects also impact target salience.

In this paper, I first introduce the general framework under which visual search research has been conducted for the better part of the last 50 years. This framework is then narrowed to prior research concerned with the three aforementioned influential factors commonly found in vision research – target-distractor similarity (TDS), target-background similarity (TBS), and background complexity (BC). I then describe Experiment 1, which was designed to test whether or not these factors apply as readily to the real world as they do to the more contrived stimuli traditionally used in this field of research. I continue by describing Experiments 2 and 2b, which build upon the results of Experiment 1 by examining these factors with greater scrutiny. Finally, results of Experiments 2 and 2b are discussed in the greater context of visual search.

2 Literature Review

For the better part of the last 50 years, vision researchers have used sanitized lab stimuli to tease out the nuances of how, *exactly,* we see and search the world around us. Much of this work is based on the groundbreaking work of Treisman and colleagues (e.g., Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Sato, 1990), which produced Feature Integration Theory (FIT). The theory states that everything in the visual field is represented as individual “features” until attention is directed to them, whereupon these features combine into “conjunctions.” Features, in the context of visual search, are all of the visual aspects of a thing – its size, shape, color, texture, and so on. Treisman’s idea, then, is that these features “float” in our visual field at or near their actual location in the real world, and that only by focusing attention on these things do they combine into conjunctions (referred to as “binding”).

The theory goes on to suggest the existence of two stages of visual search – a preattentive stage in which features are processed in parallel, which is followed by an attentive stage in which features are serially combined into conjunctions. A search task which can be completed with only the processing of the parallel stage is referred to as “Feature Search” (e.g., looking for something red when nothing around it is red – see *Feature Search* in Figure 1). A search task which requires the serial conjunction of features, however, is referred to as “Conjunction Search” (to use a classic example, a red X surrounded by green X’s and red O’s – see *Conjunction Search* in Figure 1).

In practical terms, this means that searching for something that sticks out of its surroundings is easy and takes little time, whereas finding an object that blends in takes more time and scrutiny. Certainly the search for the red X is easier in the Feature Search image over the Conjunction Search image (Fig. 1), and this difference is due once again to saliency. The red X is more salient in the Feature Search image because of the context in which it appears; one has to do little more than glance at it to know whether the red X is there and where it is located. However, when multiple objects share the feature ‘red,’ as in the Conjunction Search image, the search becomes harder,

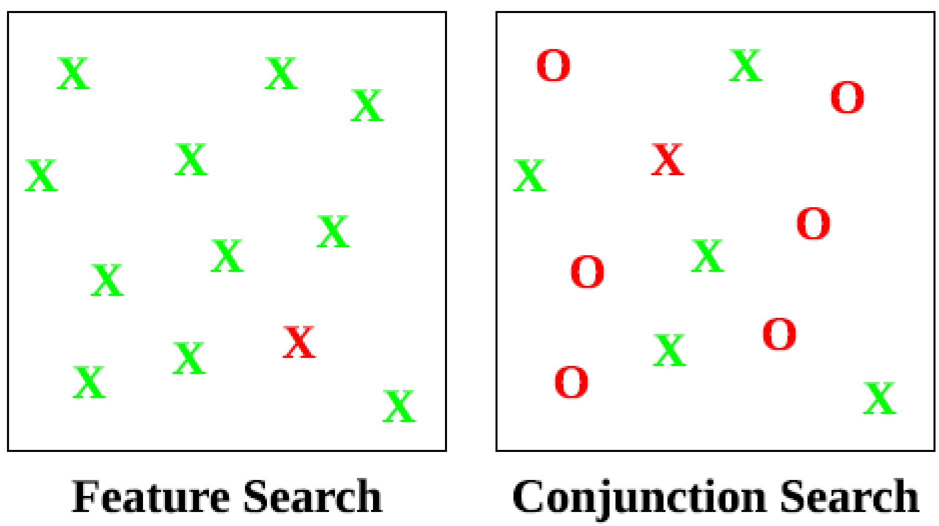


Figure 1: Examples of Feature and Conjunction Search

Feature search can be performed almost instantly, no matter how many other things are in the display (referred to as “set size”); the graph of response time by set size in feature search is a flat line. In conjunction search, however, response time slowly increases as set size increases. Importantly, the slope of the line for target-absent trials (i.e. where the target you’re searching for is absent) is about double the slope of target-present trials. This type of response pattern is the gold standard by which observations in visual search are measured.

and one must spend time examining candidate targets to find the red X hidden among them. This is because Conjunction Search requires that attention be focused on objects – in this case, X’s and O’s to determine their redness and X-ness – to combine them.

A modification of FIT was proposed by Wolfe and colleagues (e.g., Wolfe, Cave, & Franzel, 1989; Wolfe 1994a; Wolfe 2007). The idea of the new theory was (and is) that the parallel stage and the following serial stage are not independent of each other as such, but that the “results” of the parallel stage *guide* the serial search. This model was, fittingly, named “Guided Search” (GS).

To highlight the difference between FIT and Guided Search, it is useful to revisit the example of the red X surrounded by green X’s and red O’s. If the *parallel* search for the red X fails, which is to say that the red X cannot be picked out by a single feature like redness or… X-ness... (an important point on which both FIT and GS agree), then under the FIT framework search will continue through all available stimuli which fit at least one of the features, serially rejecting each in turn until the target is found. That is to say, both the green X’s and the red O’s will be searched through until the red X is found. By contrast, under the GS framework, serial search does not continue in quite the same way. Rather, the parallel stage is suggested to reject a large portion of the stimuli that do not match based on one feature while searching through those that fit another – in this example, all green X’s might be rejected while red targets are serially searched for the X (see Figure 2). This leads, therefore, to an increase in search efficiency, because the serial stage has only to look through the red targets to find the red X instead of looking through all of the green targets as well.

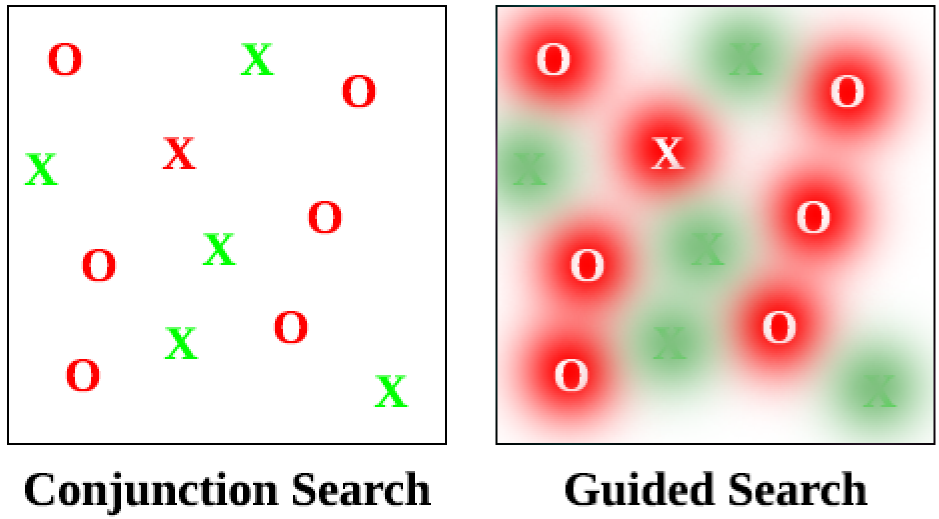


Figure 2: Conjunction Search and a Visualization of Guided Search

*FIT's conjunction search as it compares to a visualization of Guided Search. In the FIT conjunction search paradigm, **all** targets which share one or more features with the search target are serially searched. By contrast, in the GS search paradigm, targets are preattentively "screened out" by one of the features while the other feature is investigated. In this visualization, the green X's (which share the feature of X-ness with the red X) are screened out so that all red targets may be searched.*

Another account of visual search emerged at around the same time as Guided Search. Proposed by Duncan and Humphreys (Duncan & Humphreys 1989; Duncan & Humphreys 1992), the Attentional Engagement Theory (AET) attempts to account for the differences in feature and conjunction search in a different manner. Under the assumptions of AET, feature search and conjunction search are not distinct processes, but rather are two ends of a spectrum of visual search. The apparent differences between feature and conjunction search are explained by way of target-nontarget (TN) similarity and nontarget-nontarget (NN) similarity. Search is shown to be more difficult as TN similarity increases, and to be easier as NN similarity increases. That is to say, if the target is very similar to the things around it, it will be harder to find, and if the nontargets are more similar to each other, the actual target will be easier to find. They call this "spreading suppression," and in principle it is supported by Farmer and Taylor (1980), who found that increased background uniformity in a simple color-search task led to decreased search times.

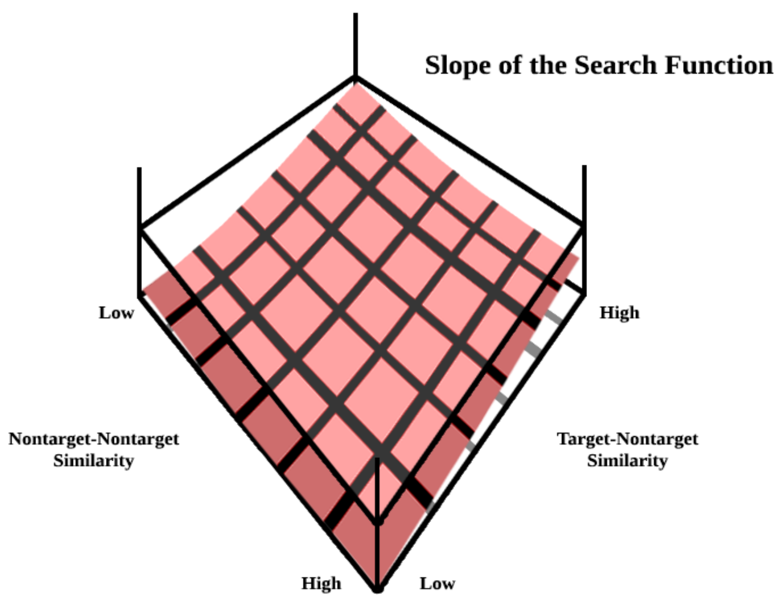


Figure 3: Adaptation of Duncan and *Humphreys' Search Surface (1989)*

Again, this intuitively makes sense, but it appears to contrast quite starkly with the model of conjunction search proposed by Feature Integration Theory. Treisman (1991) presents a defense of the Feature Integration model across four experiments, in which it is convincingly shown that the distinction between feature and conjunction search matters even when controlling for TN and NN similarity. However, Duncan and

Humphreys (1992) suggests that Treisman's approach focuses too much on the search surface (see Figure 3), which they state is merely a good summary of previous research and is not the crux of their theory. Rather, they articulate that their model is primarily concerned with attentional deployment during visual search, describing a process of selection and attentional weighting (spreading suppression) that is similar in principle to the guidance proposed by Guided Search; targets and nontargets are weighted in parallel based on their similarity to each other after the parallel stage of search, and this input informs the following stage of search. New data are presented following the specifications of Treisman (1991), and the AET model is adjusted to put more emphasis on spreading suppression. However, Treisman (1992) rejects the idea that feature binding does not play a role in search with one more experiment, testing the predictions of FIT on a conjunction search task with increased numbers of distractors which, according to Duncan and Humphreys (1992), should be affected by spreading suppression. However, the graph of RT vs set size lines up almost exactly with the predictions of FIT, showing conclusively that spreading suppression does not play a role in search for conjunctions.

Despite Treisman's heavy critique of the AET model, it may still be useful to think of search in terms of the search surface. Indeed, Treisman (1991) did not suggest that the search surface was invalid, but rather that the search surface alone could not account for the time spent during conjunction search. It is worth noting, then, that Duncan and Humphreys' search surface (1989), being composed of TN similarity and NN similarity, maps somewhat nicely onto the three search-impacting factors that are prevalent in the visual search literature, i.e., target-distractor similarity, target-background similarity, and background complexity.

Nontarget-Nontarget Similarity is conceptually almost identical to Background Complexity. If the background is less complex, it could be understood as being more similar to itself. Furthermore, by the search surface model, distractors that are similar to each other and not to the search target will fade into the background. The other measure, Target-Nontarget similarity, can be seen as a combination of Target-Distractor Similarity and Target-Background Similarity. Distractors and the background are both certainly “nontargets,” and Duncan and Humphreys (1989, 1992) might suggest that, rather than being distinct phenomena, the apparent effects of TDS and TBS on search time are really two ends of a spectrum.

Suppose you have a red sedan that you’ve parked it in a busy parking lot, and now imagine a situation in which somebody is looking for your car – perhaps you’ve forgotten where you parked, or you’ve sent your friend out with the keys to bring it around front. In this situation, what counts as a distractor, and what counts as the background? It seems obvious that the roads and buildings nearby would probably constitute the background, and that any other red vehicles probably constitute distractors, but it’s less clear how to classify a grey sedan, or a grey motorcycle. If there’s a billboard with a red sedan on it, is that a distractor, or is it part of the background? If a building in the background is simply red but isn’t shaped like a sedan, would that constitute a distractor?

The answer to the question of “what constitutes a distractor versus a background object” is not as easy to tease apart as it seems. The fundamental question is at what point these objects become just a part of the background instead of distractors, and the answer is not immediately clear. I might argue that these things are all distractors to varying degrees, based on a definition of saliency that takes into account top-down influences (is it a vehicle?) and bottom-up influences (is it red?). The real difference between distractors and background may even be a matter of Gestalt figure-ground differences, in which what things constitute distractors and which constitute background is a matter of perceptual grouping and thus, in this context, a matter of what your target is (Rubin, 2001).

Rather than speculate further on the issue of distractors and the background as distinct or as a spectrum, I will instead attempt to contextualize these possibilities as I further review the literature and my own experiments. Central to this investigative review will be the following questions:

1. How do the findings on TDS, TBS, and BC transfer from sanitized lab stimuli to real-world stimuli, if they transfer at all?
2. Are TDS and TBS meaningfully different, or are they part of a spectrum?
3. Perhaps most importantly: Can highly controlled laboratory experimentation give us an accurate view of the much more complex world in which we live?

2.1 Attempts at a Naturalistic Approach

Although these models have been hugely influential, they seem to be lacking when it comes to more naturalistic stimuli. That is to say that, while this wealth of previous research is extremely robust and often able to tease out tiny nuances of visual search mechanisms, it is not clear whether the effects will carry over when visual search is studied in a more naturalistic environment.

Let us begin with an example. Lavie and Cox (1997) describes an experiment in which six hexagonally arranged letters are shown with a congruent or incongruent distractor on one side or the other for 100ms. The six hexagonal letters are manipulated between trials to be either easy (e.g., five O’s when searching for an X) or hard (e.g., a collection of H’s and N’s while searching for an X). Curiously, when the task is easy, the type of distractor can vary response time by 20-30ms, but not when the task is hard. The effect found was statistically significant, and has theoretical implications for how attention is deployed when it seemingly is not as necessary.

Despite the theoretical implications, however, it is worth asking what the practical implications are. Consider that the effect found by Lavie and Cox (1997) is 20-30ms. This time frame is extremely short, and it would be extremely easy for noise from a more-naturalistic study to cover up or even negate such an effect. The question must then be asked: can such a nuanced finding be replicated in a naturalistic environment? To be clear, Lavie and Cox (1997) is not a bad study; it is based on sound science and has an interesting result. However, it demonstrates a theme in vision research – studies are often not replicated within a realistic environment. The findings of studies like Lavie and Cox (1997) are sound, and undeniably reveal with nuanced precision truths about human vision. However, when such tight controls are put on experiments, it is no longer clear that these findings are meaningful outside of the lab.

There are, of course, those who have used naturalistic methodologies. One such study was conducted by Cathcart, Doll, and Schmieder (1989), in a follow-up study based on Schmieder and Weathersby (1983). In it, they investigated visual search in what they have termed “urban clutter,” as the previous experiment had investigated “rural clutter.” Participants located targets within a scene containing both urban clutter (e.g., houses, cars) and rural clutter (e.g., trees, bushes). Curiously, they found that the type of clutter appeared to matter – urban clutter impaired search to a lesser degree than did the rural clutter. This experiment indicates that there may be performance differences between different types of backgrounds; however, their imagery is not the most convincing, consisting largely of computer generated, wire-frame polygons. Other limitations are self-noted by the authors, including the small sample of each type of clutter, and the difference in frameworks used by the two studies. Still, the results are not to be ignored; taken together, Cathcart, Doll, and Schmieder (1989) and Schmieder and Weathersby (1983) indicate that there may be more to be learned from the naturalistic approach. Indeed, with typical lab stimuli, it is hard to make meaningful clutter with which to test this apparently semantic effect on response time. What combination of X’s and O’s, for example, might make a convincing urban environment? How might one combine them to create a forest scene? It is perhaps the case that sanitized stimuli cannot be used to differentiate semantic differences in clutter. And, if that is indeed the case, then what other limitations of this approach might exist?

The notion of “clutter” in a more general sense is a common theme in naturalistic visual search studies. For example, Ho, Scialfa, Caird, and Graw (2001) investigated the effect of visual clutter on search for traffic signs. Subjective ratings of clutter were gathered across several images of traffic scenes, taking place in both day and night. This was followed by a search experiment using an independent participant pool who were of similar age, wherein it was found that the images rated for high clutter were somewhat predictive of search response times. This finding is supported by Donderi (2006), who notably also found that compressed file size correlates nicely with subjectively rated clutter.

Neider and Zelinsky (2011) observed similar results to Ho et al (2001) and Donderi (2006). While attempting to create systematic, naturalistic stimuli based on urban and rural clutter, they found that the most predictive element of response time was subjectively rated “clutter,” rather than the urban/rural quality of the imagery. They note, however, that while this correlational data is very strong, it isn’t as strong as correlations typically found between in more lab-controlled tasks (e.g., Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989) between RT and the number of items on screen (set size).

While compressed file size and subjective ratings seem to be reasonable estimations of clutter, two additional measures were proposed by Rosenholtz, Li, and

Nakano (2007) – Feature Congestion and Subband Entropy. These, along with Edge Density, were examined, and it was found that all three measures correlate well with previous clutter-search studies (citing, e.g., Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). It is further suggested that the Feature Congestion model may be superior because it outperforms the other two measures in the metric of color variability.

Other naturalistic studies have avoided the investigation of clutter and complexity entirely, instead investigating specific real world issues. Hollingworth, Williams, and Henderson (2001) used line drawings of real world scenes to investigate how visual information is gathered during scene viewing. They found that memory plays a role in scene viewing, and more specifically that changes could be detected in previously viewed objects. This study represents one small but confident step toward the real world from the lab.

Wolfe (1994b) tackled the issue in a different fashion, creating tiled stimuli which were meant to simulate an aerial view of rivers and lakes. In the task, participants searched for a blue lake object among rivers and off-colored lakes (representing pollution). The findings of this study mirror the expectations put forth by Guided Search (Wolfe, Cave, & Franzel, 1989; Wolfe 1994a), and it is thus asserted that the findings of basic lab studies do in fact apply to the real world. However, while the imagery used in this study and in Hollingworth, Williams, and Henderson (2001) are certainly closer approximations of real-world imagery, they remain over-simplified by comparison.

Sareen, Ehinger, and Wolfe (2016) created actual real-world imagery for use in a change blindness task. Images were created with a change both inside of a room and another change either outside of a window or in a mirror. Participants were shown pairs of imagery and attempted to detect a change between the two; however, none of the experimental manipulations were significantly predictive of RT variance between images. This perhaps suggests that there is not enough control between images, implying a difficulty with creating “lab-grade” imagery for experimentation.

2.2 Additional Control in Naturalistic Vision Research

While previous attempts at a more naturalistic approach to visual search have had some success, the gap between lab knowledge and real world knowledge is far from bridged. This, I suggest, is because those moving in the naturalistic direction have by and large failed to account for what makes the real world different from the sanitized stimuli – namely, the “real world” aspect. Experiments seemingly cover every possible facet of visual search in the lab, implementing controls upon controls to make sure nothing interferes with the effect being measured. This precision is excellent for establishing first principles and for investigating the nuances of visual search; often, results are within 100 ms each other, and sometimes significant differences are a fraction of that. Without the level of experimental control typical in vision research, the different aspects can quickly overrun each other with confounds.

Sometimes, though, we can learn interesting things from attempting to grapple with complexity. Memory research was greatly aided by the early attempts of

Ebbinghaus (1913) to make quantified measurements of memory span and training, even though he was only able to run experiments on himself. Closer to home, change blindness research has made important contributions to driving safety (e.g., Caird, Edwards, Creaser, & Horrey, 2005; Galpin, Underwood, & Crundall, 2009) and criminal law (e.g., Davies & Hine, 2007; Nelson, Laney, Fowler, Knowles, Davis, & Loftus, 2011) by getting out of the lab and into the real world.

In the specific realm of visual search, naturalistic studies have been narrowly focused on applied research that is very narrow in scope. For example, there is a wealth of information relating to how the TSA searches through baggage (e.g., Wolfe et al, 2013; Biggs, Cain, Clark, Darling, & Mitroff, 2013; Menneer, Cave, & Donnelly, 2009) and how radiologists search for abnormalities in X-ray images (e.g., Krupinski, 2005; Bird, Wallace, & Yankaskas, 1992; Kundel & LaFollette Jr, 1972), largely focusing on better detection of threats in these scenarios (Wolfe & Horowitz, 2007; Wolfe & Van Wert, 2010). These studies, while excellent uses of our knowledge from the lab, also highlight a disconnect from the laboratory data: whereas a sanitized search task of, e.g., X's and O's typically last only a few seconds, a real world a search task like searching an X-ray image can take several seconds or even several minutes (Wolfe, 2010).

This, then, is the obvious and unanswered question – will the wealth of data already collected still hold water in the real world? It is already known that there are difficulties present when attempting to take lab findings and directly apply them to the real world (Clark, Cain, Adamo, & Mitroff, 2012). It is therefore worth investigating if the most fundamental and prevalent principles learned in the lab – the effects of

Background Complexity, Target-Background Similarity, and Target-Distractor Similarity – will themselves lose anything in translation. Fortunately, each of these three factors is covered by its own extensive body of literature; this robust background provides a quite suitable basis upon which to investigate whether or not laboratory findings, *in principle*, can be trusted to apply to the real world. Should it be the case that these factors all apply, the "hurdles" discussed by Clark et al (2012) can be taken as just that – hurdles. If, however, these fundamental findings do not translate well into naturalistic settings, it may be worthwhile to evaluate other lab-based findings in visual search with more scrutiny as we try to bridge the gap to the natural world. For the sake of clarity, I will discuss each of these factors in series.

2.2.1 Background Complexity

It is perhaps strange that the background of an image, insofar as the background is understood to be distinct from distractors, would have much effect on a single-target search; one might be inclined to think that only the number of candidate targets should have an effect on search. However, background could perhaps be understood as a sort of camouflage. If one drops something small on carpet, it is noticeably harder to find if the carpet is highly patterned than if it is a uniform color.

And, indeed, we see the same thing in the lab; in even the simplest of search tasks, the uniformity of the background appears to exert a strong influence on visual search. Farmer and Taylor (1980) tested this effect by having participants look for a particular colored square in a 3x5 grid of colored squares and measuring response time.

They observed that response time was longer on trials where the background was scrambled rather than uniform lines, as well as on trials where the background colors were closer in lightness/darkness to each other. Furthermore, they noted markedly longer response times on target-absent trials. This study provides strong evidence the content of the background is important in the rapid rejection of unimportant items, allowing for such things as pop-out search. However, it can also be seen to have an impact on target-present trials, lengthening response times by as much as ~100ms in one condition and as little as ~10ms in two others, following a curve.

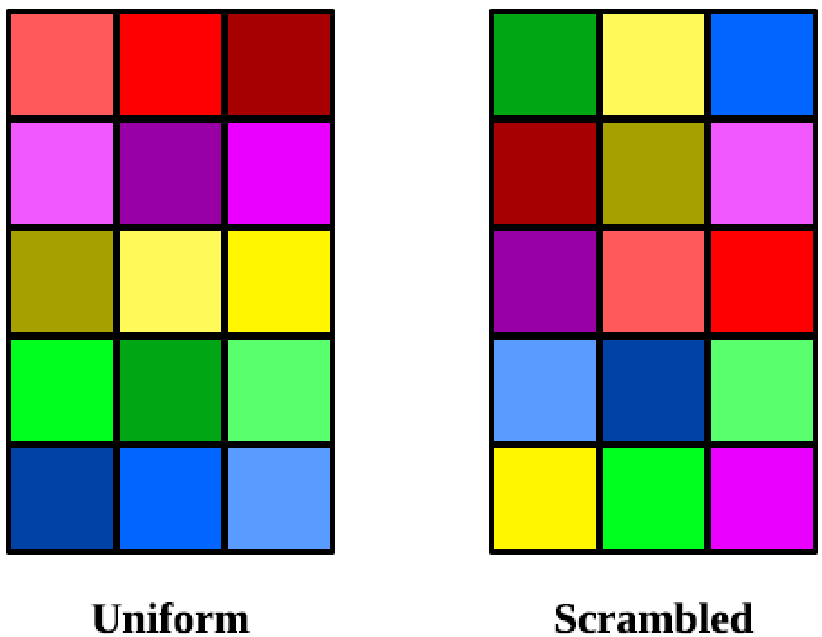


Figure 4: Example Adaptation of Backgrounds of Farmer and Taylor (1980)

The Uniform background has the three shades of each color randomized within the same row, whereas the Scrambled background randomizes the 15 tiles across the whole scene.

The argument could be made that Farmer and Taylor weren't really measuring background uniformity but rather *distractor* uniformity, as the entire visual scene is comprised of squares of different colors. Farmer and Taylor assert in this piece that the distractors are really just the squares of the same color as the target; however, the distinction is not as clear, and the question of what makes a distractor versus a background is again worth considering here. If one is searching for, e.g., the middle-hued red square, perhaps the other red squares are like the other red cars in the parking lot analogy, and the other non-red squares are like non-red cars. Are the other cars distractors in this case? Background? The distinction is hard to make, but I would argue they are somewhere in between. It is worth noting that Farmer and Taylor assert a stronger effect of background uniformity than Duncan and Humphreys (1989); this disparity can be seen as one piece of evidence that distractors and background are really just more- or less-salient nontargets, respectively.

It is worth noting that background complexity necessarily incorporates elements of background uniformity. Thus, if an experiment were to demonstrate an effect of background complexity, this would necessarily indicate that background uniformity could show its own effect. However, background uniformity may be harder to tease out in a naturalistic setting than background complexity.

Moving on, Lavie and Cox (1997) investigated the efficiency of distractor rejection during simple and complex search tasks. They showed participants six letters in a ring formation, one of which was the target. A compatible, incompatible, or neutral distractor was sometimes shown to the side. Interestingly, during the simpler tasks, the incompatible distractor had a small effect on RT, but not during the more complex tasks. This suggests that irrelevant items (perhaps interpreted as the background) can lengthen search times during simpler search tasks (e.g., during single-target search), as the low perceptual load of these tasks allows for the processing of the distractor when it otherwise wouldn't have entered perception.

As mentioned earlier, Attentional Engagement Theory provides an account of background complexity, as well. While it has been shown that background complexity and target-background similarity (TBS) cannot entirely account for the differences between feature and conjunction search (Treisman, 1991; Treisman, 1992), it must be noted that and TBS can account for a large amount of the variance between the two search times (Duncan & Humphreys, 1989; Duncan & Humphreys, 1992). Pertinently to background complexity, the search surface is very shallow as the background becomes increasingly non-uniform *except when TBS is high*. That is to say, the content of the background appears to impact search efficiency, but the effect is only meaningful when the target is more similar to the background.

2.2.2 Target-Background Similarity

Background Complexity can easily be conflated with Target-Background Similarity as simply "the effect of the background," but this broad brush would paint the nuances of the specific effects of the background. The studies on AET presented by Duncan and Humphreys (1989;1992) are a good example of this; not only do they distinguish between the two, they outline when each is pertinent and how the two effects interact.

More specifically, they note how the effects of the background are largely muted except when the similarity of the background to the target is high. Similarly, Farmer and Taylor (1980) – who investigated the effect of different colors and arrangements of colored tiles in a 3x5 grid – found that backgrounds that were more similar to the target resulted in significantly longer RTs for those trials. It is therefore important not only to distinguish between background complexity and target-background similarity, but also to investigate these effects independently.

Indeed, a study by Wolfe, Oliva, Horowitz, Butcher, and Bompas (2002) examines the effect of background “complexity” across six experiments, where complexity is measured by manipulating the similarity of the background to the target and distractors. In each of these, participants searched for a T among L’s while the background was manipulated in several ways. In one experiment, a computer-generated desk was used, and clutter (e.g., books, sticky notes) was manipulated; in another, T’s and L’s were on a grid, and the grid lines were terminated at various points to create T or X junctions. While it was observed that increasing the similarity of the background to the target and distractors significantly increased search times, the experimenters didn’t control for the uniformity of the background. Thus, it is difficult to discern if a uniformity effect might have interacted with the similarity effect.

Fryklund (1975), in an experiment in which participants searched through a 5x5 grid for five red letters in various positions and patterns. The background included distractor letters in varying positions to measure the effects of both the arrangement of distractors as well as the effects of their similarity to the target, and it was found that both of these effects were significant. It was observed that performance on a the partial-report task increased as target-background similarity decreased. While this study does not deal directly with RT, it lends further credibility to the idea that target-background similarity can impact correct perception of the target.

Similarly, a study by Neider and Zelinsky (2006a) investigated the impact of scene context on response time and gaze time. That is to say, they investigated whether or not the context of the scene could give participants cues as to where to find search targets. Using a faux-realistic background of a computer-generated desert scene, participants searched for blimps, helicopters, and jeeps in either “scene-consistent” (e.g., jeeps only on the ground) or “scene-inconsistent” (e.g., jeeps in the sky or on the ground with roughly equal probability). It was observed that RT was about 200ms faster for the scene-consistent trials, indicating that information from the scene background was used to search the scene. Similar observations were made by Henderson, Weeks, and Hollingworth (1999), who used line drawings to simulate naturalistic imagery without adding too much detail. While these studies are not strictly concerned with target-background similarity, they do suggest that the context provided by the background guides search. Thus, if one were to investigate target-background similarity, one should either implement an independent measure of scene context or, perhaps more feasibly, control for it (perhaps by flipping the image upside down). As top-down influences such as this are outside the scope of this thesis, I have taken the latter approach in my experiments.

A second study was conducted by Neider and Zelinsky (2006b) investigating the effects of target-background similarity more directly. Participants searched for children's toys on a background of tiles sampled from the target or on a blank background. For the former, the tiles were 20x20 squares taken from part of the target; a second experiment expanded this to include 15-pixel and 35-pixel tiles. This manipulation resulted a significantly higher error rate for the 35-pixel condition, suggesting TBS makes it more difficult to find the search target. This was also reflected in the response times, with response time increasing markedly with TBS. There is some question of whether or not the tiling of the background created a uniformity effect due to patterning; if it did, however, it appears to have been massively overpowered by TBS, which is consistent with the predictions of the search surface (Duncan & Humphreys 1989; 1992).

A similar strategy was employed by Boot, Neider, and Kramer (2009) and Neider, Boot, and Kramer (2010). Participants again searched for a children's toy on a tiled background made from the target. This was done for three sessions to train participants to perform the task properly, and then participants were shown novel stimuli in the same format. It was observed that the training transferred nearly perfectly to the novel stimuli, but also that training beyond about five exemplars did not further increase performance when detecting novel stimuli. Thus, it can be reasonably concluded that a small amount of training is necessary when investigating TBS, and likely other search effects. This also consistent with the findings of Neider and Zelinsky (2006b).

2.2.3 Target-Distractor Similarity

Target-Distractor similarity is similar in concept to Target-Background Similarity, however, TDS deals with salient, potential targets rather than objects which should blend into the background. Whether or not there is a meaningful difference here is unclear; Wolfe et al (2002) would suggest that TBS and TDS are really just two ends of a spectrum, as would Neider and Zelinsky (2006b). Nevertheless, TBS and TDS are often investigated a separate entities, with researchers tending to shy away from what is probably a middle ground between the two.

Through one mechanism or another, Treisman and colleagues (e.g., Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Sato, 1990), Wolfe & colleagues (e.g., Wolfe, Cave, & Franzel, 1989; Wolfe 1994a; Wolfe 2007), and Duncan and Humphreys (1989; 1992) all suggest that distractors which share similarities with the target will lengthen search times. However, the response time waters can be easily muddied by other effects (e.g., positioning, lightness/darkness, and global-local effects). Thus, it has been necessary to investigate TDS under several different conditions to confirm that, indeed, target-distractor similarity has an effect on search.

Treisman (1982) investigated part of this puzzle, looking into the effect of perceptual grouping on response time. This was tested by means of a six-object search task, which tested for positional effects as well as for an effect between parallel and serial search. In the fourth experiment, participants primarily looked for a red O or blue H in groups of red O’s and blue X’s, and it was found that the conjunction targets were hardest to find if they were on the border between the groups which each contained one of its two features. This implies that distractors being closer to a target can inhibit the detection of that target by camouflaging it, forcing serial distinction.

A similar effect was found in Driver, McLeod, and Dienes (1992). Participants searched for an X moving along one diagonal (top left to bottom right and back again) of the screen among Os moving along the other diagonal. In it, difficulty of conjunction search was manipulating by altering the phase of the X and O groups, and it was found that search became extremely difficult only when both groups were out of phase (i.e., when out of phase, half of that group would move “up” while the other half moved “down”), but not when only one group was out of phase. Duncan (1995) argues that this is evidence of a bias toward what he calls “common fate;” movement in phase allowed for perceptual grouping of the objects moving together, and thus all could be accepted as serial search targets or rejected as distractors all at once.

Building on this, Halverson and Hornof (2004) examined the effect of sparse and dense visual grouping on response times. They used displays of word lists of five or ten words which mapped onto size 18 or size 9 font, respectively. While the experiment conflates text size with sparseness, it makes the important observation that participants will tend to search sparse groups before they will search dense groups, and that these groups are searched more quickly than their dense counterparts. This is consistent with the findings of Treisman (1982) and Duncan (1995).

Thus, if distractors can camouflage a target by virtue of proximity to the target, it stands to reason that the distractors have some sort of advantage over the background when it comes to impairing search times, which is consistent with the predictions of AET (Duncan & Humphreys, 1989; 1992). This advantage is not always present; as previously discussed, Lavie and Cox (1997) found that dissimilar distractors had a small impairing effect (~20ms) on search, but only when the search was otherwise very easy (e.g., feature/pop-out search). However, it should be noted that Lavie and Cox (1997) is not inconsistent with literature suggesting that TDS has an effect.

Scialfa, Esau, and Joffe (1998) investigated the effect of target-distractor similarity on response time across younger and older groups. Participants looked for a segmented circle among other segmented circles; similarity was measured by means of rotating the distractor circles 30, 60, and 90 degrees from the target. It was observed that high TDS is associated with a significantly higher error rate, and high TDS was similarly associated with higher response times. The simplicity of this experiment is reminiscent of Farmer and Taylor (1980), which by virtue of simplicity established both background content and TBS as significant effects in visual search. Similarly, the findings of Scialfa, Esau, and Joffe (1998) convincingly show that Target-Distractor Similarity has an impairment effect on response time.

The next natural question is the mechanism by which TDS impairs RT, and the magnitude thereof. In what is termed the Frankenbear experiment, Alexander and Zelinsky (2012) investigate TDS when parts of the target are transplanted onto distractors (e.g., the head of the target teddy bear appears on a distractor teddy bear as well as the target). Two relevant findings were reported: subjective similarity scores significantly predicted the number of parts transplanted from the target to the distractor (i.e., a distractor sharing more parts with the target was deemed more similar), and both RT and accuracy were impaired on increased similarity trials. These findings are consistent with previous findings that TDS negatively impacts search.

2.3 Summary

This literature review has demonstrated the recurring nature of Target-Distractor Similarity, Target-Background Similarity, and Background Complexity in the visual search literature. These factors are all well established as impacting visual search in a variety of contexts, and as such they can be considered fundamental factors.

Given the robust nature of these three factors, they are appropriate for translation to the real world. If they translate well to naturalistic contexts, it can be reasonably asserted that other, more nuanced effects found in the more-sanitized laboratory experiments will also translate. If they do not translate well, it will indicate that we should reconsider the understanding of visual search that has so far been derived from laboratory environments. I have hypothesized in between these two extremes: it is my suspicion that these fundamental factors will translate quite well to the real world, but that there will likely be some errata that arise in the natural world that otherwise have not been observed in more traditional experimentation, similar to the “hurdles” observed by Clark et al (2012).

In Experiment 1, I attempted to demonstrate that the effects of TDS, TBS, and BC would translate to the real world, and predictably certain problems arose. In addition, certain unexpected effects emerged between the three images tested in the experiment. Thus, Experiment 2 was devised to better control for the problems found in Experiment 1 as well as to investigate the apparent errata that arose. I will discuss these experiments, the aforementioned problems, the unexpected effects, and the following experiments in much greater detail in the following chapters. To reiterate, the primary questions with which I am concerned are as follows:

1. How do the findings on TDS, TBS, and BC transfer from sanitized lab stimuli to real-world stimuli, if they transfer at all?
2. Are distractors and background meaningfully different, or are they part of a spectrum?
3. Can highly controlled laboratory experimentation give us an accurate view of the much more complex world in which we live?

3 Experiment 1

In Experiment 1, I investigated single-target search under 5 conditions for each of 3 images. Participants performed a task similar to *Where's Waldo*, searching for a small target image within a larger, more complex image. To the extent that results are found to be inconsistent with prevailing theories of visual search, this would indicate that there may be some discrepancy between traditional visual search and more naturalistic visual search, consistent with the conclusions of Wolfe (1994b) and Wolfe (2010).

3.1 Method

3.1.1 Participants & Data Trimming

Seventy undergraduate students were recruited through the subject pool at Michigan Technological University and tested for visual acuity using a Snellen eye chart and for colorblindness using Ishihara's Tests (Ishihara, 1980). Five participants failed to pass one or both of these tests, and their data were excluded. One participant's data were excluded when it was found they did not use the chin rest. An additional 12/70 participants' data were excluded because they their data followed a steep accuracy drop-off below 80% accuracy; it is suspected these participants did not do the task correctly. See Appendix I for additional details. In sum, 18 out of 70 participants were excluded from Experiment 1, or just over 25%. This is quite a high rate of data attrition, and steps taken to mitigate this rate for Experiment 2 will be discussed in that chapter.

3.1.2 Materials

The experiment was programmed in and performed using the Psychology Experiment Building Language (PEBL) (Mueller & Piper, 2014). Participants sat at a desk using a chin rest to moderate visual distance from the computer screen. Chin-rest-to-screen

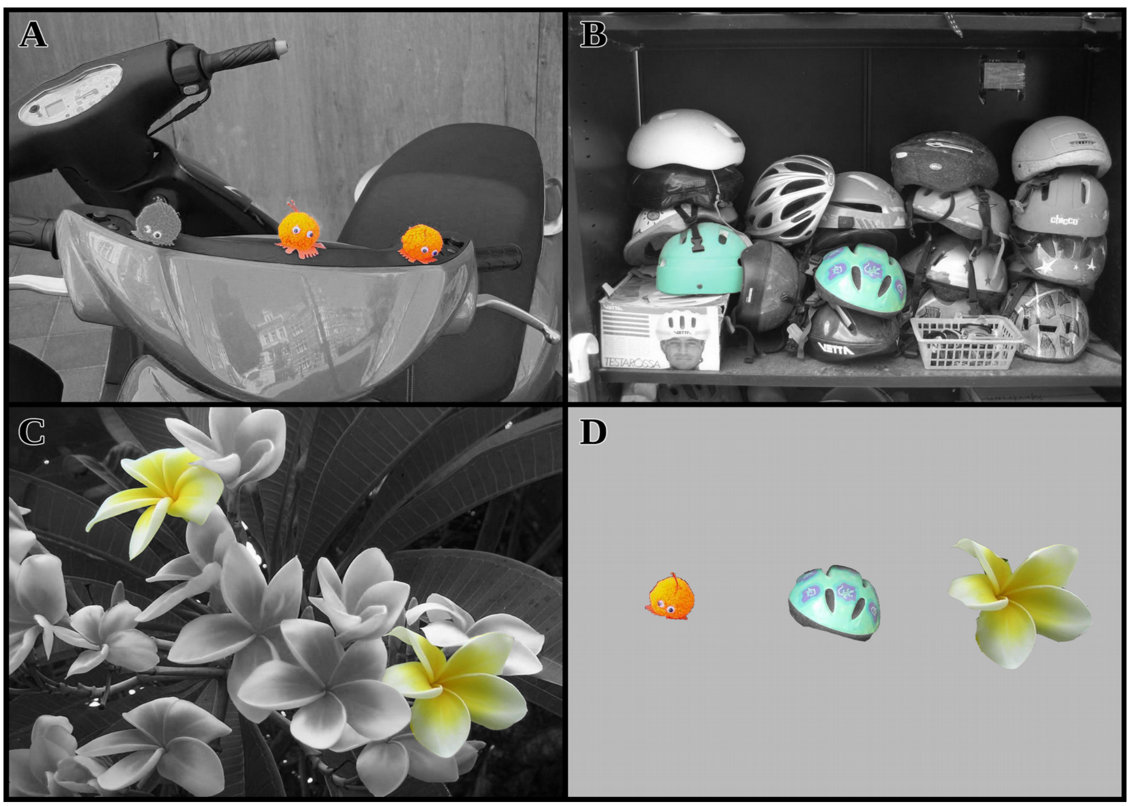


Figure 5: Experiment 1 – Example Imagery and Stimuli

Images A-C are examples of images the participants might see. In each, two of three possible stimuli are highlighted to facilitate search for only those objects. Image D contains one possible target for each image. For a full list of stimuli and imagery, see Appendix II.

distance was 47cm (18in), and the screen had a resolution of 1920 x 1080px with a diagonal measurement of 54.6 cm (21.5 in). Stimuli ranged from 83 x 78px (1.26 degrees of visual angle) to 292 x 194px (8.83 degrees of visual angle).

Three images were selected from the MIT 300 saliency benchmark image set (Judd, Durand, & Torralba 2012), and three stimuli that were determined visually similar by myself and my lab group were selected from each. When the selected stimuli were visually similar except in color, they were color shifted to increase similarity (e.g., the helmet on the left in Fig 5, Image B was originally orange). The large images were converted to greyscale, while the selected stimuli remained in color; then, for each trial, the colored stimuli were overlaid on the greyscale image according to condition. For each image, each of the three targets was a target in five unique conditions, creating 15 unique trials per image (45 trials across all three images, following a 3 x 3 x 5 within subjects design). These “stimulus conditions” were as follows:

1. Target present, set size = 1
2. Target present, set size = 2
3. Target absent, set size = 0
4. Target absent, set size = 1
5. Target absent, set size = 2

Key interactions are condition 1 vs 4 (present vs absent, set size 1), condition 2 vs 5 (present vs absent, set size 2, to establish slope), and condition 1 vs 3 (present vs absent, varying set size). Target-absent trials were managed by way of greying out the stimulus, and participants were told that if a target was grey, it was to be considered absent. Two practice trials were administered to ensure participants understood how to perform the task. A target-present, two distractors condition was considered but not implemented; participants saw each image multiple times, and could feasibly have known instantly that the target was present if all three stimuli were colored. For each trial in the second condition (target present, one distractor), the distractor stimulus was randomly selected from the two non-target stimuli.

3.1.3 Procedure

The experiment followed a 2 (present or absent) x 3 (set size 0, 1, or 2) x 3 (base images – motorcycle, flower, or helmets) within-subjects design, with the exception of set size 0 target-present trials, as these are logically impossible. Each trial consisted of first observing a target stimulus for any amount of time, and participants pressed the spacebar when ready to begin the trial. The target was then masked for 800 ms, after which the greyscale image would appear. Colored stimuli were simultaneously overlaid on the greyscale image based on stimulus condition (see Appendix II for further explanation). Participants would press the spacebar when they either found the target or determined its absence, and the whole image would become greyscale if it wasn't already. Participants would then either click the location where they found the target, or would click a button labeled "Absent." In this manner, I assessed accuracy as well as response time.

3.2 Results

This experiment was conducted to address three specific questions: whether or not TDS, TBS, and BC, broadly speaking, transfer well into the real world; whether or not TBS and TDS are meaningfully different; and, in the bigger picture, whether or not highly controlled experimentation in the lab can give us a truly accurate view of the natural world.

3.2.1 Expected Results

TDS, TBS, and BC are well established as fundamental factors in visual search; as such we would expect these effects to persist across many types of naturalistic search tasks, as they have across many lab experiments. In order to make meaningful conclusions about these three factors, though, we must start at the most basic level of scientific analysis – simple calculations of response time as it is affected by set size, presence or absence, and, in this case, accuracy.

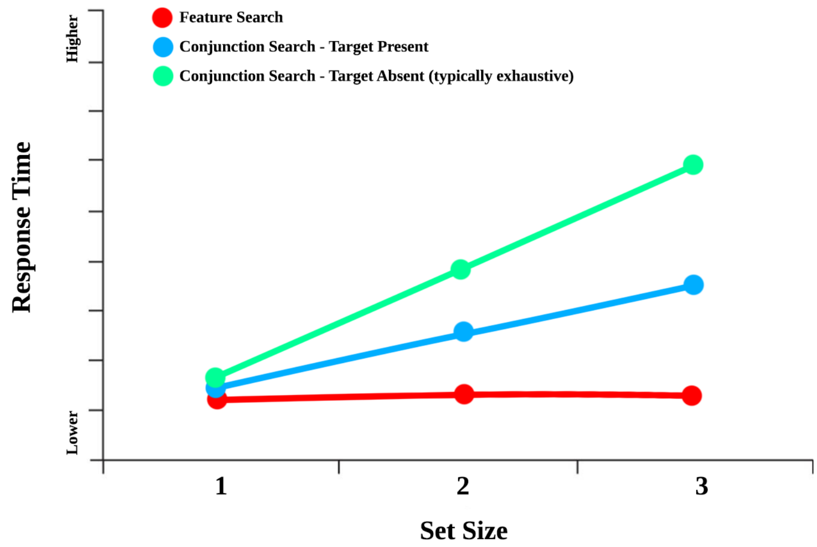


Figure 6: Feature Search and Conjunction Search

In feature search, the zero-slope pattern persists regardless of set size, with a minor drop in accuracy at higher set sizes. By contrast, conjunction search will slow as a result of both set size and target absence. In conjunction search, slope of target-absent search times is double that of target-present search.

Response time has proved to be the gold standard in visual search tasks, while accuracy is not consistently a concern of visual search researchers. However, in typical visual search tasks, it is very easy to tell whether or not you have found your target, and for this reason it is possible that in a more naturalistic setting accuracy might suffer, because stimuli could conceivably be harder to recognize than a simple colored letter or symbol; thus, information about accuracy was also collected.

The set size by present/absent effects and interaction effect on response time are especially important, as they demonstrate the most fundamental assumptions of all three of the standard models of visual search. These assumptions are as follows:

- In a feature-search task (remember back to the red X among green X's) should take a set amount of time, regardless of set size (see Figure 7); the exclusionary feature makes the target "pop-out" from the background/distractors; this effect is what I aimed to test with conditions 1 and 3. This expected search pattern is sometimes referred to as "pop-out search." While this experiment was originally intended to measure feature search, it did not accomplish this goal.
- In a conjunction search task (e.g., the red X among green X's and red O's), one would expect that response time would increase as a function of set size, and that target-present trials will be faster than target-absent trials. Additionally, for target-present conjunction search, the slope of the graph of response time by set size should be about half of the slope of the same graph for target-absent trials (see Figure 7). The conjunction search pattern is what I aimed to test by comparing conditions 1 and 2 with conditions 4 and 5, respectively. These search patterns are reflective of what is termed "serial, self-terminating search;" the targets are searched one at a time until the target is found. Logically, conjunction search when the target is absent will typically be exhaustive, which is to say that all relevant targets are searched before search is terminated.

Finally, if, in a naturalistic task, accuracy is high (> 90%), it can be safely assumed that accuracy is not an issue when translating to the natural world. However, sufficiently lower accuracy than a typical visual search task would indicate that these tasks are harder, for one reason or another, than a typical lab experiment on visual search. It is hard to make an error when searching for a red X; can the same be said for visual search in naturalistic imagery?

These effects are predicted by all prevailing models of visual search. Thus, if these criteria are met in a simple experiment such as this one, evidence is provided for the easy translation of TDS, TBS, and BC to naturalistic imagery, and the theories under which these effects emerge (FIT, GS, AET) will be validated as predictive of naturalistic settings. If any of these criteria are not met in such an experiment, however, it may indicate that our current understanding of visual search is incomplete, and that naturalistic imagery contains complexities of visual search that are not apparent in traditional experiments.

3.2.2 Results

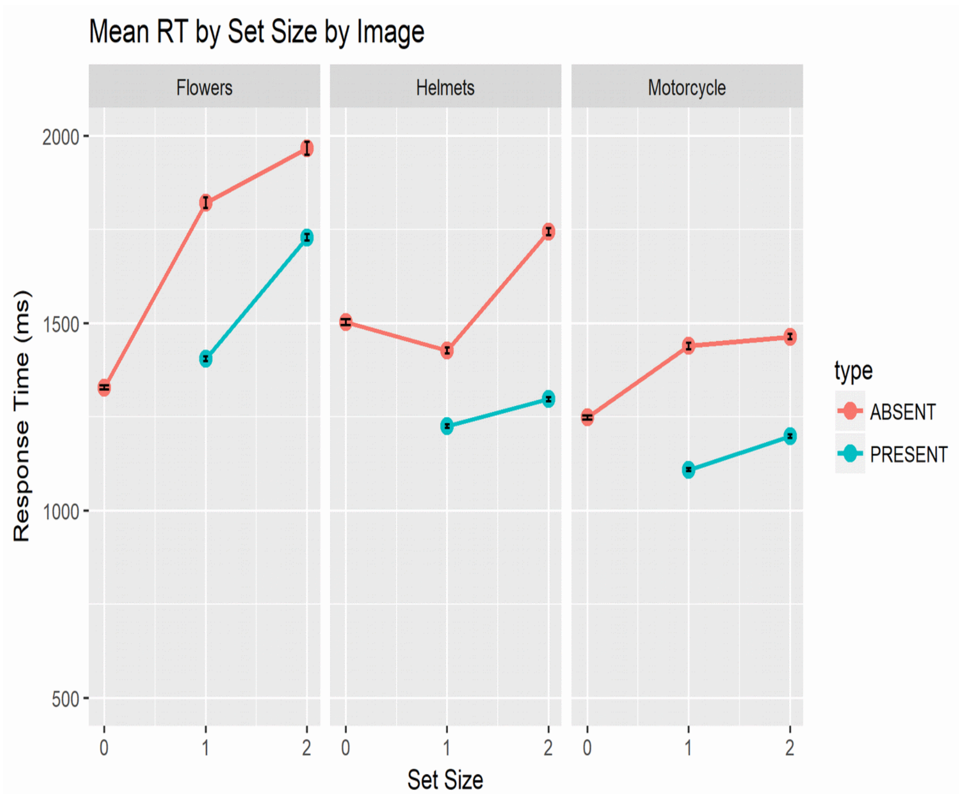


Figure 7: Experiment 1 – Response Times by Set Size by Image Type

The three separate graphs represent individual results by image type. This was initially investigated because graphs of overall response time for what should have been a relatively simple task did not line up with expectations. Error bars reflect standard error of the mean.

Overall accuracy was 80.9% across all participants; after exclusion of the 12 participants who were suspected of not doing the task correctly, overall accuracy across participants rose to 89.7%.

Results of Experiment 1 provide evidence that our fundamental understanding of visual search is roughly accurate, but may be incomplete. A factorial ANOVA model of response time, on only accurate trials, treating participant number as a randomized factor, revealed reliable main effects of set size, $F(2,98)=15.96$, $p < 0.001$, and present/absent, $F(1,48)=7.88$, $p = 0.007$, as well as a reliable main effect of base image, $F(2,101)=24.148$, $p < 0.001$. There was no interaction effect found between set size and present/absent, $F(1,51)=0.051$, $p=0.82$. The interaction between set size and base image was marginally significant, $F(6,98)=2.084$, $p=0.062$, indicating that the two effects may have been found to interact if the experiment had more power. Finally, the set size by present/absent by base image three-way interaction was not significant, $F(2,77)=2.0$, $p=0.14$; however, the evidence against an effect is lacking. As two of these results were only weakly rejected, I thought it prudent to investigate these as possible effects in Experiment 2.

3.3 Discussion

The results of Experiment 1 sometimes conform to expectations and sometimes do not. In this section, I first discuss the implications of specific results. I then identify potential problems in the experiment that could explain these results, and I conclude with a discussion of possible broad explanations for the unexpected results.

3.3.1 Implications

To begin, it is worth revisiting our basic expectations for search, vis-a-vis Figure 6. In feature search (attempted in condition 1 [target present, set size=1]), we would expect to see a flat response, regardless of set size. I would argue that condition 1 did not accomplish this goal, as in this task the participant must both determine whether or not an object fitting the description of the target is present *and* confirm the target as correct before responding; by comparison, a real feature search task would be faster, requiring only preattentive information (e.g., it's the right color and it's there or not) to make the present/absent decision. However, condition 1 can still be compared to condition 4 (and condition 2 compared to condition 5) 4 to test conjunction search.

Conjunction search makes more predictions about search, and this is where most of the consistencies and inconsistencies with current theories lie. In a conjunction search, one would expect that target-present trials would be faster than target-absent trials, and one would also expect that higher set size leads to increased response time. Furthermore, these two effects should interact in a specific way: the slope of target-absent RT should be roughly double the slope of target-present RT for conjunction search (see Figure 6). The results presented here are consistent with the first two effects, and thus the predictions of FIT, GS, and AET; a main effect of set size and present/absent were both observed, indicating that core assumptions of these theories apply very strongly to the real world. However, these terms did not interact in this experiment, and in fact a quick glance at Figure 7 shows three completely different effects where this interaction should be apparent. This suggests that our current theories of search may not have the full picture of conjunction search – a well-supported phenomenon in the lab.

Another phenomenon in these results is the main effect of base image and its weak interaction with set size. While it cannot be asserted that the current theories of search do *not* predict differences between imagery (indeed, taken to its logical extreme, this seems an obvious conclusion to draw), the models do not necessarily predict the interaction. That is to say, current theories of search can account for a difference between images; however, the fact that the image modulates the effect of set size on RT is not necessarily something that would be expected based on previous research. In fairness, this effect is weak, and merits further testing before making a strong assertion about its implications.

Finally, while there are main effects of set size, present/absent, and base image, as well as a marginal interaction of base image with set size, there is no set size by present/absent by base image interaction, though this was only marginally insignificant. Taken with the marginal interaction of base image and set size, this result suggests that there may be more to learn about our search process; however, the evidence is weak at best, and therefore the only strong conclusion that can be made is that more data are needed to draw a strong conclusion. Thus, I investigated these two effects with additional power in Experiment 2.

These results suggest that our understanding of visual search could be incomplete, and that translation from the sanitized lab setting to the real world may not be as easy as one might have hoped. However, there are a great many potential problems with this experiment, as well as several good hypotheses that may explain the inconsistencies with prevailing theories of search. These are laid out in series, and methods for testing or controlling for each are discussed. Then, I discuss the specific implications of the results found.

3.3.2 Potential Problems

A distinct learning effect was observed across the first ten trials or so for each participant, regardless of base image. While the basic results of this experiment persist with the inclusion of these learning trials, a simple solution to this problem is to have each participant view each image only once. Within-image reliability would drastically drop in this case, but this was avoided by using visually similar base images to create categories of images (e.g., two participants might each see a different-but-visually-similar image of marbles), and by extension creating between-image reliability. This is important; in traditional experiments, reliability is a non-issue due to the simple nature of the imagery, whereas in naturalistic imagery I suspect there may be differences across images that might not appear in more sanitized imagery. Additionally, participants were trained much more generously in Experiment 2, helping to eliminate any learning effect resulting from lack of familiarity with the task.

Several potential problems exist within the stimuli alone as well. For example, he rightmost and leftmost stimuli in the helmet image may have been too tattered to distinguish without directed attention, as they are both obscured by the buckle from the helmets above. This might have reduced participants' ability to recognize the helmet when it was a target, resulting in the lengthened response times compared to the central helmet. However, the central helmet had odd markings on it, which could have contributed to the increased response time for that image; the markings are purple on the teal helmet, but the other two helmets were simply teal.

Finally, the experiment was somewhat underpowered. After excluding roughly 18/70 participants (~25%) for simply not doing the task correctly in one form or another, experimental power was too low to make call these conclusions reliable.

3.3.3 Potential Explanations

In addition to these potential problems, there are several potential explanations for the divergence from FIT/GS/AET predictions on measures of present/absent by set size interaction and the differences between images. Should these explanations hold up, it would indicate that there is no divergence from these theories despite the appearance of divergence.

Below, I present four hypotheses for why there appear to be differences across images and why there does not seem to be a present/absent by set size interaction. Because these images are naturalistic and therefore may contain additional complexity not accounted for by current theories of visual search, it should be noted that these hypotheses are not mutually exclusive; indeed, it should not be surprising if all four hypotheses prove relevant. For this reason, orthogonal measures of each were developed for Experiment 2. These hypotheses may be familiar by now:

1. **Background Complexity –** In short, the background “junk” in the images may be having an effect on search. There is a trend of increased response time for image backgrounds which seem to me to be more complex in nature. The motorcycle image has just one object in it, upon which the three possible targets sit; the helmet image is a little less uniform, with several greyed out objects in

the background that are arranged somewhat uniformly; and the flower image is perhaps the least uniform, with many greyed out objects scattered across the image with seemingly no rhyme or reason. This is not to be confused with the Target-Background Similarity hypothesis; the Background Complexity hypothesis is only concerned with complexity, not similarity.

2. **Target-Background Similarity –** In contrast to the Background Complexity hypothesis, the basis of the Target-Background Similarity hypothesis is that the similarity of the background to the search target is the cause of the variance seen across images. For example, it is probably not as difficult to distinguish an orange blob toy from the motorcycle upon which it sits as it is to distinguish a flower from a field of flowers or a helmet from a stack of helmets. Ergo, under this hypothesis, the background complexity only matters insofar as the target is similar to the background.

3. **Target-Distractor Similarity –** Also based on the concept of similarity, the Target-Distractor Similarity hypothesis is the idea that the similarity of the distractors to the target is impacting the results. In the motorcycle image, one can see three similar orange blobs, but with their faces facing different directions. The motorcycle image is arguably further along this spectrum, with the outer helmets bearing some resemblance to each other in both shape and the manner in which they are obscured by helmets above them. The flower image had especially long response times, even for this task in which overly long

response times were observed; perhaps the flowers are especially difficult to distinguish from one another.

4. **Memory –** It is entirely possible that, due to the repetitive nature of the experiment, participants developed a memory for the stimuli over time (in fact, several mentioned it after the experiment).
5. **Image Idiosyncrasies –** Though it is possible the differences between images are simply a result of the previous three hypotheses, it is also possible that the three hypotheses will not adequately explain the difference. Prudence dictates investigation of this possible phenomenon.

It should be noted that, while the initial intent of this experiment was to measure feature search in naturalistic imagery, this is not in practice what was measured; rather, the imagery consisted largely of low-set-size conjunction search. I would argue that it might not be a feasible to translate proper feature search into the real world. It is rare that we know for certain that, e.g., the only pair of keys on the counter are, in fact, ours; instead, one must focus attention on the keys to verify them as our own. At a crowded event, one cannot not know for sure that they placed the only blue coat upon the coat rack; rather the blue coat must be inspected to verify it as one's own. It is therefore argued that single-target search may be more appropriate for a naturalistic approach than true pop-out search, and thus Experiment 2 investigated search in a similar context to Experiment 1 with additional control measures implemented.

3.4 Summary of Experiment 1

In this chapter, I discussed the motivation for Experiment 1, its methodology, expected results, and actual results. Of particular interest, I observed certain results that were consistent with current theories of visual search, suggesting that these theories can be applied to the real world. However, I also observed certain inconsistent results, which suggest that these theories may need to be modified to give a more complete view of the real world.

Certain fundamental questions remain. With some effects that are consistent with FIT/GS/AET and others that are not, will these results persist in more tightly controlled conditions? Are there truly differences between imagery in a simple search task, or did that effect merely pop up due to low power? These questions are investigated in Experiment 2, in which I attempt to control each of the previously discussed hypotheses in an orthogonal manner while investigating these curious effects.

4 Experiment 2

Building on the findings of Experiment 1, a new experiment was conducted. Additional controls were implemented in order to investigate single-target search in naturalistic imagery. Participants again performed a task similar to *Where's Waldo*, searching for a small target image within a larger image. Additional control was exerted over the image-creation process so that the hypotheses could be assessed orthogonally. Image creation, methods, results, and implications are discussed.

4.1 Addressing Previous Issues

Several potential problems were previously outlined with Experiment 1 – among them, a learning effect, low experimental power, and the presence of a difference in response time between images. All of these were considered when creating Experiment 2.

Dispatching of the learning effect was a simple matter. First, participants were run through the task from Experiment 1 with experimenter oversight and explanation to make sure that they understood the task and could competently carry it out. Second, Experiment 2 was made to fall more in line with typical visual search experiments, showing each image only twice rather than 15 times. The two trials were distinct, particularly because the target to be searched for was present in one trial and absent in the other. For the absent trial, the target was designed to look visually similar to the target of the present trial.

From Halverson and Hornof (2004), Treisman (1982), and Lavie and Cox (1997), it can be reasonably concluded that a target hidden near other distractors would be found slower than a target outside of such a grouping. As such, care was taken when designing stimuli to avoid creating clusters of distractors, instead spreading them out through the whole image. Furthermore, additional care was taken in the creation of images with more than one candidate target so that one was not obviously different from the others. No targets were obscured by background targets. Targets and candidate targets were altered so that they were not obscured by the background (only necessary in the Lego image set). It should be noted that target and distractor locations were not specifically controlled for other than making conscious design decisions to avoid perceptual grouping described in Treisman (1982).

Finally, experimental power posed a potential problem of logistics; several steps were taken to mitigate this problem. Primarily, more participants were recruited – 120 instead of 45. 120 participants were recruited because power analysis of the results of the previous experiment suggested the need for twice as many people, and the extra 30 account for the data attrition rate observed in Experiment 1. Obviously the hope was to solve the issues leading to such high attrition, but the 120 number served as a safety net in the event I could not. Cohen’s d for several measures in Experiment 1 ranged from 0.27 to 0.45, and this was used as the baseline when doing the power analysis for Experiment 2.

4.2 Method

Before I discuss the typical inclusions in method sections (participants, materials, procedure, etc.), I must first take a moment to describe how the stimuli were created. Creating single-target search tasks in naturalistic imagery requires quite a lot of control, and thus a discussion of the control methodology is warranted.

4.2.1 Limitations on Stimuli Creation

Creating naturalistic scenes that are also controllable presents a unique challenge in itself, and, as shown in the preliminary experiment as well as Sareen, Ehinger, and Wolfe (2016), this endeavor is not without its pitfalls. However, some guidelines for such creations exist in the literature, albeit probably inadvertently.

Treisman (1982), Halverson and Hornof (2004), and Neider and Zelinsky (2006a) all discuss the problem of perceptual grouping. These studies vary in setting; while Treisman (1982) and Halverson and Hornof (2004) demonstrate this effect using stimuli that are typical of the lab setting, Neider and Zelinsky (2006a) use children’s toys on a tiled or blank background. In both cases, the result is the same: perceptual grouping negatively impacts search by grouping the target with the distractors. Thus, it was reasonable to avoid this sort of grouping. In naturalistic stimuli, this proved tricky, as the typical control for this – a ring configuration – does not appear often in nature.

Biederman, Mezzanotte, and Rabinowitz (1982) investigated scene perception in the context of violated expectations. Participants were shown a line-drawn, naturalistic scene for 150ms, and it was found that semantic violations (e.g., a sofa larger than it ought to be and/or floating in the sky) were detected in this brief period. The implication here is that, in a single glance, participants can tell what isn't as it should be in a visual scene. With regard to stimuli creation, it meant that "naturalistic stimuli" could not simply be real or real-looking objects; they must also occur in a natural and expected context. Image manipulations that appeared unnatural were thus avoided.

4.2.2 Additional Control Measures in Stimuli Creation

Stimuli were created across 8 different, paired image classes (Marbles & Legos, Peppers & Raspberries, Leaves & Locks, Sunflowers & Coins) to account for possible differences between imagery as well as for he anticipated possibility of top-down differences between organic and man-made stimuli. In each set, four visually similar images (based on photographic elements such as lighting, viewing angle, etc.) were gathered from the public domain or creative-commons-licensed sources. From these four images, targets were extracted from the background and kept in color while the background was desaturated (turned grey), which allowed for the easy swapping of backgrounds to test the BC and TBS hypotheses. Stimuli to test TDS were created by using stimuli from the "other" image class in the pair (e.g., distractors when searching for a marble were either marbles or legos, colorized to match the target).

It is not clear whether or not distractors and background are part of a continuous spectrum or are two distinct concepts; thus, the grey background and vividly colored target/distractors control for this by making each distinct from the other.

4.2.3 Participants & Data Trimming

Based on the parameters from the power analysis, 120 participants were recruited through the subject pool at Michigan Technological University as well as from the local community. Participants were tested for visual acuity using a Snellen eye chart and for colorblindness using Ishihara's Tests for Colorblindness (Ishihara, 1980). Three participants were excluded for low accuracy (<75%), and 2 as response time outliers. Overall accuracy was 85.9% across all participants; after exclusion of the 5 aforementioned participants, overall accuracy across participants rose to 86.9%.

4.2.4 Materials

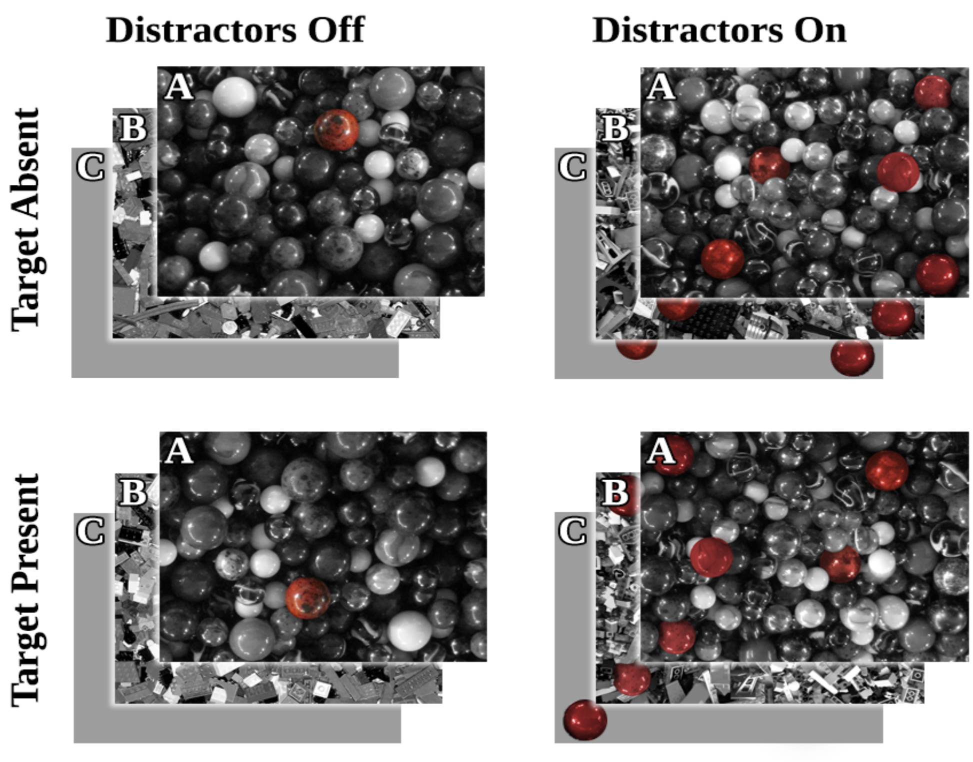


Figure 8: Example Set of 12 Images

For each of the 8 image sets, a single participant saw one of the three images in each stack. Thus, distractors and target presence/absence were counterbalanced within subjects, but backgrounds were counterbalanced between subjects.

The experiment was programmed in and performed using the Psychology Experiment Building Language (Mueller & Piper, 2014). Participants again sat at a desk with using a chin rest to moderate visual distance from the computer screen. Chin-rest-to-screen distance was 47cm (18in) for most participants; it was somewhat shorter for 23 participants. All screens had a resolution of 1920 x 1080px with a diagonal measurement of 56.4cm (21.5in).Stimuli ranged from 53 x 48px (1.52 degrees of visual angle) to 171 x 146px (5.41 degrees of visual angle).

Images were gathered from public sources with licensing allowing for modification and non-commercial redistribution (see Appendices III and IV). Four visually similar images for each of 8 categories were gathered. From each of these four images, two more were created with differing backgrounds – one swapped with a background from its partner image class, and one a blur of color, making 12 total images (see Figure 6). I will refer to these as the Base, Swap(ped), and Blur(red) backgrounds, respectively. Additionally, for image A in both Distractors On stacks, another image with different distractors (taken from the partnered image class) was created. Thus, each image class (marbles, legos, etc.) has a total of 14 images across 12 image types (3 [backgrounds] x 2 [distractors on/off] x 2 [target present/absent], + 2 [swapped distractors for two images per set]). Backgrounds remained grey in an attempt to draw a clear distinction between background and distractors. For a full example image set, see Appendix III.

Each participant saw 32 unique images – 4 from each of the 8 image classes – and then another 32 images which were variations of the first 32 images in terms of background. A new target was shown for the second 32 trials (present if the previous corresponding trial was absent, and vice-versa) and the background was changed to one of the two remaining backgrounds. The three backgrounds served to test TBS and BC; the Base background compared to the Swapped background compares Target-

Background Similarity while using a fixed amount of complexity, while the Swapped vs Blurred comparison serves to test the effect of Background Complexity. Due to the method by which BC was tested, it is partially entangled with TBS. Teasing apart the specific influence of BC on its own is beyond the scope of this research.

4.2.5 Procedure

The experiment followed a 2 x 2 x 3 mixed factor design, wherein, *for each of 8 images sets*, 2 (set size) x 2 (present or absent) x 3 (background condition) were conducted within subjects, with background condition being counterbalanced between subjects. Each trial consisted of first observing the target stimulus for as long as desired. When ready, participants pressed the spacebar to begin the trial. They were then shown a mask for 800ms, after which the large image was shown. Upon finding the target or determined its absence, participants again pressed the spacebar. Finally, participants localized the target with a mouse click or, if the target was absent, clicked a button labeled “absent.” In this way, I planned to record response time as well as accuracy. Due to a coding error, the swapped distractor trials were not included in the actual experiment, and thus TDS could not be measured.

4.3 Results

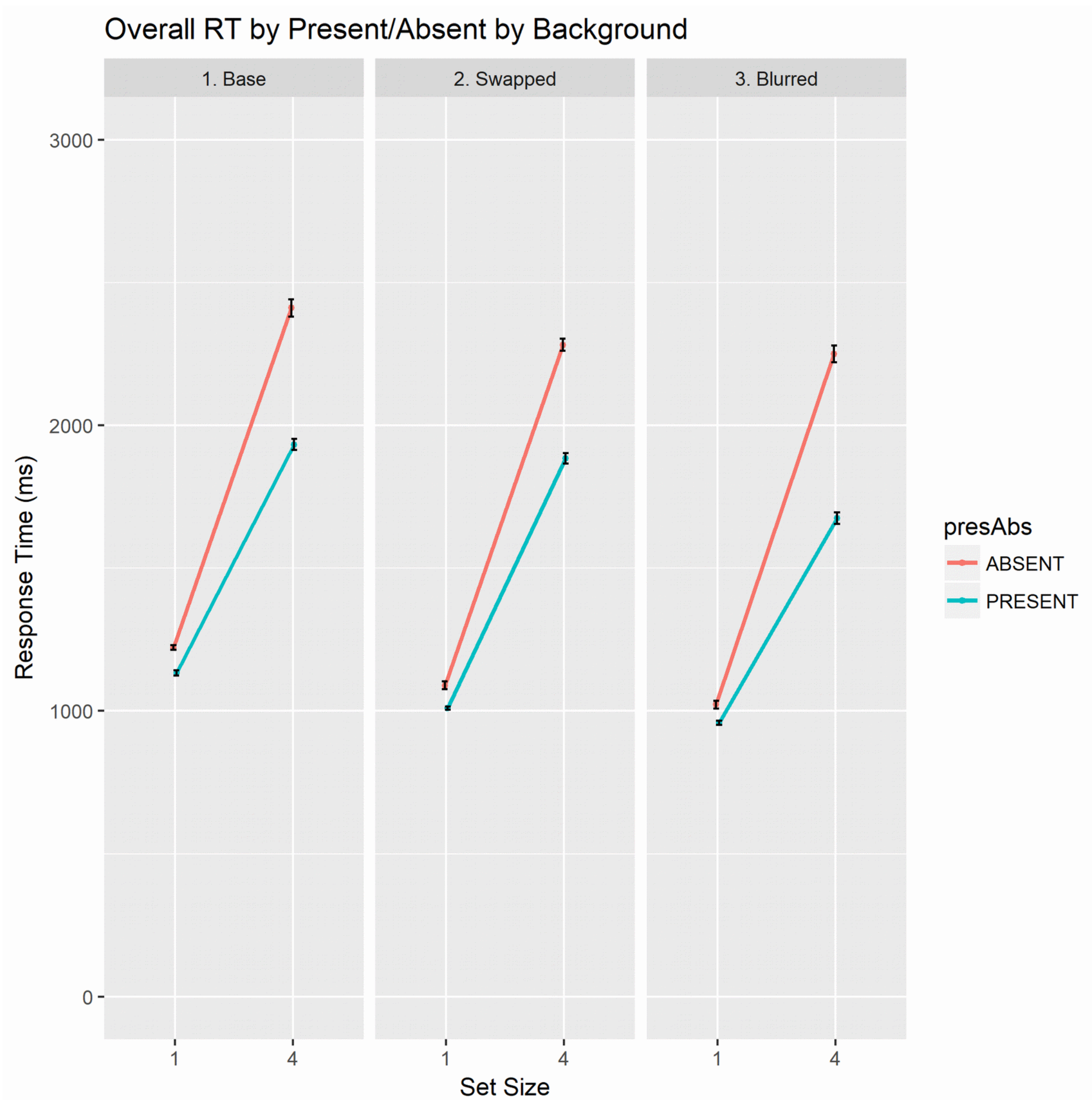


Figure 9: Overall Response Time Collapsed Across Image Classes

Error bars reflect standard error of the mean.

The results of this experiment are striking in that they at first appear to be a blatant rejection of the results of Experiment 1 (compare Figure 9 to Figure 6). However, as can be seen in Figure 10, the apparent consistencies to not necessarily persist across different types of imagery. I begin with the most fundamental results.

Once again, a factorial ANOVA model of response time treating participant number as a randomized factor was performed on only accurate trials, which revealed reliable main effects of present/absent, $F(1,103)=147$, $p<0.001$, and set size, $F(1,103)=1022$, $p<0.001$, and in this experiment we found the interaction between present/absent and set size, $F(1,103)=93.97$, $p<0.001$, that was lacking in Experiment 1. Additionally, background condition (base, swapped, or blurred) was found to have a main effect, $F(2,206)=29.53$, $p<0.001$, but did not interact with present/absent, $F(2,206)=1.73$, $p=0.18$, set size, $F(2,206)=2.17$, $p=0.12$, or the present/absent by set size interaction, $F(2,206)=1.47$, $p=0.23$, indicating that the effect of background condition is an independent effect that simply shifts the graph of other effects up or down. Thus, background condition is treated as its own effect for rest of the analysis.

When controlling for these fundamental effects, image class – i.e., marbles, or leaves, etc. – was found to have a significant main effect, $F(7,718)=44.98$, $p<0.001$. Additionally, image class shows significant interactions with present/absent, $F(7,712)=31.39$, $p<0.001$, set size, $F(7,714)=19.54$, $p<0.001$, and the present/absent by set size interaction, $F(7,721)=18.89$, $p<0.001$, indicating that the image class modulates both set size and present/absent response times as well as how these two main effects interact.

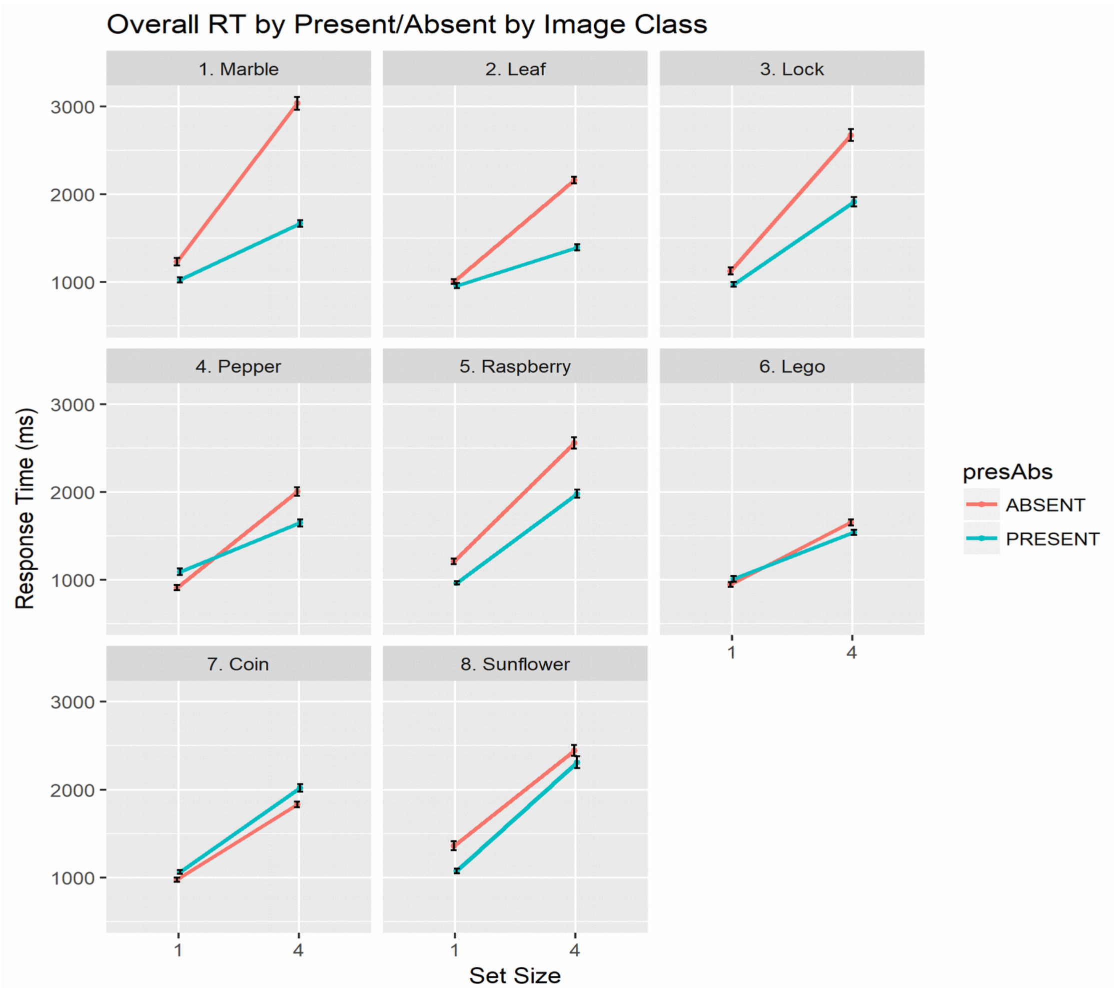


Figure 10: Experiment 2 – RT by Image Class, Set Size, and Present/Absent

The images above clearly demonstrate significant differences between image classes. It should be noted, however, that these differences collapse into much more uniform results when measured across imagery (see Figure 9). Images are arranged in descending order of difference between line slopes. Error bars reflect standard error of the mean.

Additional modeling of the background condition (wherein the condition not concerned was subset out of the data before modeling) revealed main effects of all three background pairings – Base vs Swap, $F(1,103)=11.33$, $p=0.001$, Swap vs Blur

$F(1,103)=17.58$, $p<0.001$, and Base vs Blur, $F(1,103)=68.17$, $p<0.001$. It should again be noted that background condition did *not* interact with any other main effects.

Finally, accuracy by image class did not correlate with the magnitude of the set-size by present/absent interaction by image class, $r=0.167$, $df=6$, $p=0.693$. This indicates that the main effect of image class on response time is not simply a function of participants' ability to accurately identify the target in different image classes.

4.4 Discussion

First and foremost, it is appropriate to revisit the consistencies and violations of basic principles of the prevailing theories of visual search. To reiterate, all current theories of search assert that present trials take longer than absent trials; that larger set size has an impact on conjunction search but not feature search (Duncan and Humphreys might dispute the mechanism by which this occurs, but I assert that they would not dispute the findings themselves); and that the present/absent and set size effects interact such that slope for target-absent trials is twice that of target-present trials for conjunction search tasks.

These most fundamental results of this study are good news for the prevailing theories of visual search. Because present/absent, set size, and the interaction between the two are all significant effects (and, I assert, strikingly similar to a typical graph of response time by set size and present/absent – compare to Figure 6), the primary assumptions of these theories are validated by Experiment 2. This is perhaps expected; if graphs of response time typically look like this across several different experiments, then, even if those experiments are largely sanitized lab experiments, one would expect them to transfer in a fairly robust manner. The real question is of how these effects might change when translated to the natural world.

This experiment provides some answers to this question. It was found that background condition had a small (compared to other effects – the F statistic is quite robust), non-interactive effect on response time, indicating that the background is likely entirely independent of the typical set size and present/absent effects. The difference between the base and swapped backgrounds (important for the target-background similarity hypothesis) is about 100ms, and similarly the difference between the swapped and blurred backgrounds was also around 100ms. All of this suggests three things about the prevailing theories of visual search: that the points on which they disagree are fairly minor; that the points on which they do agree are robust; and that the points they don't discuss, such as differences across image classes, is quite important.

This has certain implications for our current understanding of search, and there are two reasonable conclusions that can be drawn from this finding. The first is that target-background similarity might not have the interactive effect with background content that is predicted by AET. AET predicts that background content and target-background similarity are both small effects on their own (with content being considerably smaller on its own), but that they have a synergistic effect that makes them greater than the sum of their parts when both are high. This effect is not apparent in the results presented here; rather, the relationship between the two appears to be quite linear. This perhaps warrants additional testing, as the TN similarity difference between base and swapped backgrounds could feasibly be different than the NN similarity difference between swapped and blurred backgrounds. However, this additional nuanced testing is beyond the scope of this research.

On a more basic level, the effect of the background and its independence from other effects perhaps indicates something fundamental about search that may have eluded traditional lab studies. This is a process that does not impact the shape of the response time graph, but does affect how high or low the general shape is. Further, the lack of interaction with set size suggests that this is a one-time effect rather than a persistent effect; that is to say, the effect does not reappear as one shifts one's gaze around the same visual scene. These conclusions, in consideration with the finding that the background also does not interact with present/absent, suggests that the background has an effect on some preattentive, perhaps-parallel process; it serves as a one-time attentive "shock," if you will, impacting response time once as the image is processed for the first time, but not on subsequent fixations.

The effects of image class are perhaps the most interesting result. Not only does image class show a main effect on response time – suggesting that it, in and of itself, can impact RT – it can be further observed that image class strongly interacts with all of the fundamental effects on search predicted by the prevailing models, including set size, present/absent, and the set size by present/absent interaction. This is important, maybe even the most important result of this paper: while it is true that the three major models of search would probably predict (or at least, would not contest) the finding that different types of imagery can impact RT, these theories do not necessarily predict that the subject matter of the image *can modulate* the effects of set size, present/absent, and their interaction. In more-human terms, this means that image class not only produces an effect of its own on response time, it also impacts the effect of set size (suggesting that it is easier or harder to locate new targets based on image class), the effect of present/absent (suggesting that it is easier or harder to confirm or reject targets), and perhaps most interestingly, image class impacts the synergistic effect of set size by present/absent.

The nature of this last interaction of image class is not immediately clear, but it can be made sense of in the context of Figure 10. The graphs of the eight image classes are arranged in descending order of the difference between the present and absent search slopes, and with this arrangement we can see evidence that image class could manipulate what type of search is performed. For example, the Marble image class follows standard trends of serial, self-terminating search, whereas the Lego image class follows a trend of exhaustive search, even for target-present trials. This indicates that, on an otherwise quite similar task, the particular nature of the target and its distractors (even when controlling for target-distractor similarity to a degree) can impact not just search times but the fundamental nature of how that search is performed.

Thus, it appears that, in visual search, the foreground (i.e., target and distractors) may matter more than the background for response time and predictions thereof. The foreground appears to impact all sorts of things about the nature of search; Figure 10, as well as the previous discussion of image class, support this idea. This is not to say that set size, present/absent, etc. do not matter in search – indeed, Figure 9 would likely look much different if these factors proved meaningless in the real world – but it does appear that something about the nature of the search target matters in addition to these fundamental factors. In only some of the cases, it was hard for participants to unambiguously identify the target.

This is perhaps still an issue of memory; once participants began a trial, they did not have access to a reference image, as is typical in search tasks. This is presumably because a reference image would form the perfect target-similar distractor, and visual search tasks are typically concerned with response time. It is possible, then, that because participants did not have a reference image, they needed to recheck certain locations to make sure of their answer. This could account for the apparent lack of difference between present and absent trials seen for some image classes but not others. This difference could also take the form of a sort of lack of interaction, potentially caused by participants verbally memorizing objects (i.e., memorizing by some feature of the object – the reflection of the marbles, perhaps) in some classes, and being unable to use this method in others. Such an examination could be a good topic of future research, but is beyond the scope of the current investigation.

Perhaps, then, this kind of memory issue is simply inherent to visual search, particularly in the natural world where targets are inherently more complex than a red X among green X’s and red O’s.

4.5 Summary of Experiment 2

Taken together, the results of Experiment 2 indicate that prevailing theories of visual search are quite good approximations, but are perhaps incomplete in their predictions of the natural world. There is a remarkable amount that these models correctly predict, including but not limited to the effects of set size, present/absent, and their interaction on conjunction search. Search times are considerably longer than traditional search tasks – 1000-3000ms in this task, versus well under 1000ms in traditional search tasks (e.g., Treisman and Gelade [1980] only found such large search times at much higher set sizes than I have tested here); one might have expected the extra complications inherent in messier, more-naturalistic stimuli to have a more diverse range of effects. Instead, the effects of set size and present/absent scale up to the real world quite well *on the whole (Figure 9);* it is only underneath the surface that the deviations from current theories appear (Figure 10).

This experiment has thus established background complexity and target-background similarity as unique modulators of response time, and has additionally found that image class has a modulating effect on three fundamental search effects (set size, present/absent, and the interaction between the two). While current theories predict the translation of the fundamental effects to the real world, they do not necessarily predict that the subject of the imagery should modulate these effects, particularly under only some conditions. Thus, only the translation of the effect of Target-Distractor Similarity to the real world remains.

5 Experiment 2b

As Experiment 2 suffered from a coding error which disallowed collection of data for the test of Target-Distractor Similarity, an additional, short experiment was carried out in order to collect this missing data. Methodology for Experiment 2b closely mirrored that of Experiment 2, but with a small modification to allow more efficient collection of data to test the Target-Distractor Similarity hypothesis. Specifics of methodology and results are discussed.

5.1 Method

The experiment followed a 2 x 2 x 2 mixed factor design, where 2 (set size) x 2 (present or absent) were conducted within subjects, while 2 (distractor congruence) was conducted between subjects.

5.1.1 Participants

26 participants were recruited from the local community. All participants passed tests for visual acuity using a Snellen eye chart and for colorblindness using Ishihara’s Tests for Colorblindness (Ishihara, 1980). No participants were excluded as accuracy outliers, and the same metrics discussed in Experiment 2 were used for excluding outlier RT’s.

5.1.2 Materials

The experiment was programmed in and performed using the Psychology Experiment Building Language (Mueller & Piper, 2014). Participants again sat at a desk with using a chin rest to moderate visual distance from the computer screen. Chin-rest-to-screen distance was 47cm (18in) for all participants. All screens had a resolution of 1920 x 1080px with a diagonal measurement of 56.4cm (21.5in). Stimuli ranged from 53 x 48px (1.52 degrees of visual angle) to 171 x 146px (5.41 degrees of visual angle).

Images were reused from Experiment 2, including images created for Experiment 2 but which were not used due to the aforementioned coding error; details of how these images were gathered and created did not change between experiments, and will not be revisited here.

Each participant saw 32 unique images from the base background condition described in Experiment 2, and then another 32 images in which modifications were made to key trials. For all 32 trials of the second group, a different target was searched for (if the target was present in the first trial, it would be absent in the second, and vice-versa). Key trials were those where set size was 4; for all 4-set-size trials in the second group, a different set of distractors were implemented.

Each participant saw 32 unique images – 4 from each of the 8 image classes – and then another 32 images which were variations of the first 32 images. A new target was shown for the second 32 trials (present if the previous corresponding trial was absent, and vice-versa). For trials with a set size of 4, the distractors in one trial were congruent (e.g., marble distractors when looking for a target marble) while in the other trial they were incongruent (e.g., lego distractors when looking for a target marble).

5.2 Results

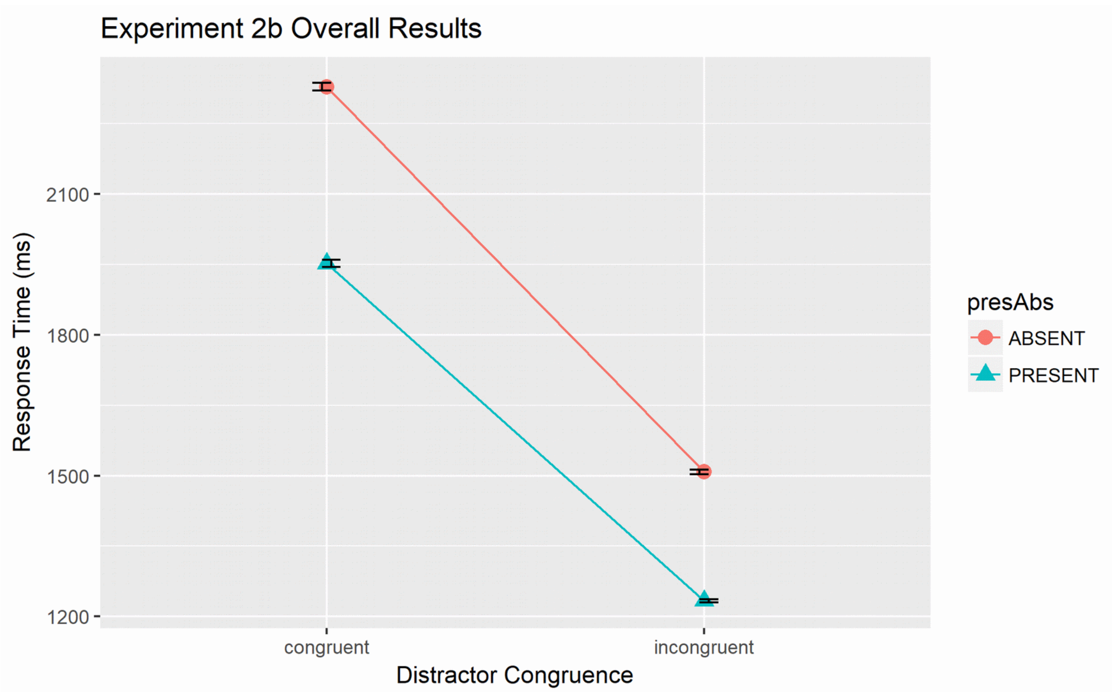


Figure 11: Experiment 2b – Response Time by Distractor Congruence

Here, we see the effect of distractor congruence on response time. A congruent distractor is one that belongs to the same image class as the target (e.g., searching for a marble among marbles. An incongruent distractor is, therefore, one that is not in the same image class as the target (e.g., searching for a marble among legos). Error bars reflect standard error of the mean.

Results of Experiment 2b in many ways confirm the findings of Experiment 2, as well as showing interactions between the newly introduced variable of distractor congruence and other variables previously established as significant in Experiment 2.

Overall accuracy was 90.1% across all participants. No participants were excluded for excessively low accuracy in this experiment. This accuracy result is 3.2 percentage points higher than in Experiment 2. Additionally, accuracy by image class once again did not correlate with the magnitude of the set-size by present/absent interaction by image class, $r=-0.60$, $df=6$, $p=0.115$. This once again indicates that the main effect of image class on response time is not simply a function of participants' ability to accurately identify the target in different image classes.

In this final experiment, I again used a factorial ANOVA model of response time, using only accurate trials and treating participants as randomized factors to avoid unjustified repeated-measures effects. To ensure the validity of Experiment 2b, I first tested for the main effects seen in Experiment 2. And, indeed, they show up again in Experiment 2b: I observed main effects of both set size, $F(1,18)=158$, $p<0.001$, and present/absent, $F(1,18)=13.91$, $p=0.002$, and the interaction between them reoccurred as well $F(1,18)=13.11$, $p<0.001$. Additionally, the main effect of image class reappeared $F(7,125)=6.755$, $p<0.001$, and the interaction effect of image class on present/absent remained significant, $F(1,125)=6.024$, $p<0.001$.

Figure 12: Experiment 2b – Response Time by Congruence and Image Class

Image classes are arranged in the same order as they were in Experiment 2 (Figure 10). Error bars reflect standard error of the mean.

Moving on, a main effect was observed for distractor congruence, $F(1,17)=90.944$, $p<0.001$, indicating that the effect of distractors changes based on whether or not they are from the same image class as the target. Furthermore, a significant interaction effect was seen between image class and distractor congruence, $F(7,125)=6.049$, $p<0.001$, indicating that image class modulated the effect of distractor congruence. Finally, a three-way interaction was observed between present/absent, image class, and distractor congruence, $F(7,125)=2.606$, $p<0.001$. This indicates that distractor congruence modulates the effect that image class exerts upon present/absent effects.

5.3 Discussion

The fact that several main effects and interactions found in Experiment 2 carried over to Experiment 2b speaks to the reliability of these effects and serves to validate the new observations in Experiment 2. This should not be surprising, however; the effects observed in Experiment 2 were massive when observed across 120 people, and it thus stands to reason that these effects, robust as they are, could very well be observed in a lower sample size.

The particular effects of concern in this transfer are present/absent, set size, and the interaction between the two, because they verify that Experiment 2b has been done correctly and that the observed effects are reliable and reproducible. Also of concern was the impact of image class, which in Experiment 2 was shown to not just impact response time on its own but also to modulate the effects and interaction of present/absent. Indeed, these effects were again observed in Experiment 2b.

The rank-ordering of image class also stays mostly consistent, with the possible exception of Locks and Peppers. Marbles still show the strongest congruency effects, while Sunflowers and Coins show the weakest. This indicates that the rank order of image classes established in Experiment 2 is reliable and reproducible.

By verifying that the results of Experiment 2 carried over to Experiment 2b (given that Experiment 2b is based on Experiment 2), it establishes that Experiment 2b is a reliable replication of Experiment 2, which is necessary to ensure that the test of distractor congruence takes place within a sufficiently similar context to the one in which it was intended. That is to say, because Experiment 2b was created to test distractor congruence, and because this test was meant to occur within the context of Experiment 2, it was prudent to ensure the test took place in a context that was as similar to Experiment 2 as possible.

Thus, we arrive at the primary observation of Experiment 2b: distractor congruence. The effect of distractor congruence can be understood as a comparison between searching for a target within similarly colored distractors from the same image class and searching for a target within similarly colored distractors from a *different* image class. Distractor congruence is important because it is a necessary part of Target-Distractor Similarity; thus, as the effect of distractor congruence on response time was highly significant, it can be reasonably concluded that Target-Distractor Similarity plays an important role in visual search in the natural world. This is again good news for current theories of visual search; FIT, GS, and AET all predict that objects which are more similar to each other will be harder to differentiate. However, what these theories perhaps would not predict is how distractor congruence *interacts* with the other observed effects and interactions.

When the three distractors are incongruent, participants seem able to go directly to the fourth, congruent remaining object to confirm or deny it as the target image. Indeed, even when the target is absent, participants are able to reject the target more quickly than if it were present with congruent distractors. This indicates that, for incongruent distractor conditions, participants can avoid serial, self-terminating search in favor of single-target search. Much like in feature search, participants do not need to check the distractors before making a yes or no decision. In the congruent condition, however, we do not see this trend; instead, typical examples of serial, self-terminating search versus exhaustive search can be observed (for target-present and target-absent conditions, respectively).

This trend is not always observed, however, and this is because of the three-way interaction between image class, present/absent, and distractor congruence. Across image classes, we see differences in this trend, consistent with the finding from Experiment 2 that image class interacts with distractor congruence. In the Marble and Leaf image classes, for example (see Figure 12), participants are seemingly able to perform single target search *regardless* of distractor congruence. In other classes (e.g., Legos), participants seem unable to do anything except exhaustive search, indicating that present/absent did not matter; it should be noted, however, that in these cases incongruent search was still faster, indicating that search probably covered at least one incongruent distractor but not all, and that this persisted regardless of present or absent.

Of particular interest is that Sunflowers and Coins show a reversed effect from what might be expected, in which present, congruent trials take longer than absent, congruent trials. This might be because present targets were very similar to congruent distractors, whereas absent targets may have been more distinguishable somehow. Regardless of the true cause of this phenomenon, it is apparent that response time is influenced by distractor congruence in addition to image class.

5.4 Summary of Experiment 2b

Experiment 2b has yielded results supporting the final hypothesis of Target-Distractor Similarity, which further vindicates the predictive power of prevailing theories of visual search in the natural world. It should again be stated that these models of search provide incredibly accurate descriptions of the average visual search process in naturalistic imagery. However, a closer look at results indicates that, while these theories are excellent predictive models overall, they do not provide a complete understanding of naturalistic visual search. Idiosyncrasies between naturalistic images, particularly regarding the foreground objects in such imagery, exert a great deal of impact over search, even against an effect as robust as distractor congruence has proved to be. If FIT, GS, and AET are to improve their predictive fit in naturalistic settings, these sorts of idiosyncrasies will need to be accounted for.

6 General Discussion

Over the course of three experiments, I investigated the transfer of common, fundamental principles of current theories of visual search to a more naturalistic setting. This primarily took the form of testing several hypotheses about factors which could impact visual search in the natural world, as well as performing tests of the fundamental principles that impact all search. After Experiment 1, additional hypotheses were incorporated in an attempt to explain results that seemed inconsistent with models of visual search.

I begin this section with a discussion of the primary hypotheses, followed by a discussion of the current theories of visual search and their fundamental findings, and I conclude with a discussion of future directions for this line of research.

6.1 Hypotheses of Visual Search

In traditional visual search experiments, three effects are commonly observed to impact visual search. They are: the effect of target-distractor similarity, the effect of target-background similarity, and the effect of background complexity. As such, it was reasonable to hypothesize that these three common effects might encompass all there is to know about search; they constituted the null hypothesis. Each of these hypotheses was supported in my experiments; however, certain other effects persisted after the variance created by these effects was accounted for.

Additional hypotheses were derived as possible explanations for the strange results – that is to say, the distinct differences between images in not just response time, but in general search patterns – found in Experiment 1. Experiment 1 of course had certain experimental problems (e.g., it was underpowered, images were repeated several times, etc.), and while these problems could feasibly have been the cause for the interesting results, it was prudent to also consider another hypothesis; as images were seen multiple times, it was possible that participants were memorizing the locations of certain stimuli and simply going to that location to confirm the target as present or absent. A fifth hypothesis was tested implicitly – the hypothesis that differences between images really did matter. If it happened that all of these hypotheses could be controlled for and the differences between images persisted, then it would be reasonable to conclude that differences across images could matter in search. Of course, as with most things, the answer turned out to be more complicated than a simple yes or no in this regard.

Through controlling for all of these hypotheses, it was found that the differences between images did in fact make a difference for search. However, this effect did not appear alone; in addition to its appearance, I observed strong effects of fundamental search processes that underlie TDS, TBS, and BC – namely, effects of set size, present/absent, and the interaction between these two, as well as an independent effect of the background, all of which are inherent to conjunction search processes that have been studied over and over in the lab.

6.2 Theories of Visual Search

Feature Integration Theory, Guided Search, and Attentional Engagement Theory all make a good approximation of how search applies to the real world. Results collapsed across image sets in both Experiment 2 (Figure 9) and Experiment 2b (Figure 11) both look strikingly like basic conjunction search. It can therefore be concluded that these theories of visual search have uncovered certain fundamental aspects about the nature of visual search; they have been tested repeatedly against highly controlled laboratory imagery, and now these effects appear once again in a much more naturalistic setting.

Despite the strength and ubiquity of the predictions of FIT, GS, and AET, it would be a mistake to conclude that these theories encompass all that there is to know about visual search. Due to the nature of Experiment 1, there was reason to perform an exploratory analysis in Experiment 2, and it was observed that, while the results as a whole collapsed into trends predicted by FIT, GS, and AET, the underlying trends were not as uniform. In both Experiment 2 and Experiment 2b, there was a significant main effect of image class, which is to say that varying the subject matter of the images in question produced radically different response time results. In a conjunction search task such as those done in Experiments 2 and 2b, where a single target must be located either on its own or amongst 3 distractors on an otherwise grey background, trends should mirror what we see in Figure 9. However, the results by image class in Experiment 2 (Figure 10) upend this assumption.

A cursory look reveals that image class impacts not just response times in all conditions, but that several different types of conjunction search are being performed. Marbles are perhaps the most standard image class; the slope of the target-absent response time line is about double that of target-present, and it provides clear evidence that, in the marble image class, serial, self-terminating search is performed all or most of the time. However, in other image sets – Legos, for example – we see a trend of exhaustive search; while set size increases response time somewhat, more interesting is that, regardless of whether or not the target lego appears in the image, participants are searching through all of the legos every time. It should be noted that these trends are what emerge from four distinct yet visually similar images; it is hard to explain this as simply conjunction search.

Results for each image class are not just distinct from each other, however; instead, they form a trend in which the set size by present/interaction changes slowly from a serial, self-terminating search to a serial, exhaustive search. This provides evidence that, in conjunction search, the phenomena of serial, self-terminating search and serial exhaustive search could lie on two ends of a range of search styles rather than being distinct from each other.

Finally, I would like to discuss Single-Target Search in the context of Figures 10 and 12. In two cases in Figure 12, the apparent effect of Single-Target Search is noted (namely, in the Leaf and Marble image classes). I will here take Single-Target Search to specifically mean the equivalent of pop-out search in naturalistic imagery.

That is to say, much like the red X among green X's, there is only one candidate; however, unlike feature search, a target marble must still be confirmed to be the correct marble, unlike the red X, whose "red X-ness" is preattentively apparent. It is not clear that Single-Target Search should be considered its own phenomenon; indeed, it could be argued that Single-Target Search is just conjunction search among weak distractors, and this is a slant I myself might be inclined to argue.

However, the existence of this apparent phenomenon brings up a much earlier question: whether or not distractors are meaningfully different from the background. It is apparent from Figure 12 that the red marbles were no more effective than red legos at deterring the observer from the true target during target-present search; however, it becomes apparent in target-absent search that they are somehow distinct, as the marbles are inspected much more closely than the legos in this scenario. Whether or not this distinction is just two closer points on a spectrum is currently unclear, and would require further investigation.

6.3 Accounting for Image Class

Image class has emerged as a dominant factor in influencing search times, but so has distractor congruence. While it is possible that the impact of the background has been muted by the grey nature, it cannot be denied that image class appears to exert an extremely large effect on response time, accounting for over 30% of the variance in Experiment 2; by contrast, the impact of the background in the same experiment was a mere 7%. Such a result would not necessarily have been predicted by current theories of visual search, and thus warrants further discussion.

Perhaps the most obvious possibility would have been accuracy; if some targets were simply harder to identify, or some images were simply harder to search through, this may have explained this effect. However, this was not the case; accuracy did not correlate with the effect of image class in either Experiment 2 or Experiment 2b. However, had this been the case, one still might not expect image classes to vary significantly; rather, one would expect to see singular images posing more or less of a problem for viewers, and a much more random assortment of image difficulty would have been found. This was not the case.

Another possibility is that of background complexity versus background uniformity. In this investigation, the decision was made to operationalize commonly observed effects of the background as background complexity, as components of background complexity (such as background uniformity) cannot be teased apart from, e.g., target-distractor similarity. Because of this, it could be argued that some essential component of background uniformity was left out. However, Duncan & Humphreys (1989; 1992) would suggest that background uniformity only matters more than a small amount if target-nontarget similarity is high. Indeed, the complex versus non-complex background yielded quite a small effect, which might be expected in Duncan and Humphreys' model when target-nontarget similarity is controlled for.

A third possibility comes from a post-hoc analysis of the stimuli. Some stimuli seem to be more “verbalizable” in terms of specific features, i.e., they may have features that can be stated to oneself to aid in detection. Perhaps, for example, the Marbles were easier to distinguish because they had reflection patterns which could be analyzed. With Leaves, perhaps the way the leaf is shaped and the direction it “points,” if you will, made them easier to search for. In the Sunflower imagery, by contrast, the Sunflowers are perhaps hard to distinguish from each other on the basis of any verbalizable feature. Similarly, in the Coin imagery, it could be argued that some potential targets would be hard to distinguish on the basis of a verbalizable feature.

This possibility starts to uncover a better explanation for the effect of image class. In Experiment 2b, the primary investigation was of distractor congruence. In this experiment, when distractors are congruent, results of Experiment 2 are replicated; however, when the distractors are incongruent, the effects observed in Experiment 2 are heavily diminished (see Figure 12). This interaction yields insight into the true causes of the apparent effect of image class. When the distractors are like the target, the typical effects of visual search are observed to scale across image classes; however, when distractors are unlike the target, these effects all but disappear. This suggests that the biggest impact on search may not be the subject matter as such, but rather the nature of the other candidate targets with which the target appears.

The question is then worth asking: what is it about this effect that could explain the gradient from serial-exhaustive search to standard conjunction search patterns seen in Figure 10? The best explanation is that distractor congruence could cause participants to recheck candidate targets, perhaps to compare two or more viable search targets before responding. This fits well with Bamber (1969)'s model of target matching in visual search, which posits that a same-different judgement must be made when performing such a search task. Two mechanisms are posited – a fast, identity-reporting mechanism, and a slower, serial-processing mechanism.

Thus, it can be seen that, for those image classes which closely resemble typical conjunction search, this may be due to having verbalizable features to aid in distinguishing a target from congruent distractors. As this ability is diminished, however – perhaps by, e.g., sunflowers that are especially hard to differentiate – it is possible that rechecking of the target and another candidate target occurs, leading to what appears to be serial-exhaustive search after the target is found and recognized. This provides a more full understanding of search in more naturalistic environments; when distractor congruence is high, we find that it modulates what might otherwise be normal, predicted search patterns in a rather extreme way.

6.4 Future Directions for Research

Overall, the current understanding of visual search is quite good; the current theories of search map quite well onto the "real world," as one tends to refer to it. However, as has been demonstrated here, there is much we do not yet understand about the real world. When one attempts to translate lab findings to more naturalistic environments, the findings predict several overall trends found in the natural world, but break down when asked to account for the idiosyncrasies that come from such an environment.

While this can perhaps be accounted for by distractor congruence, several questions remain that are beyond the scope of this research. Chief among them is how specifically distractor congruence modulates search patterns, and whether or not distractor congruence maps onto target-distractor similarity in a meaningful way. I might argue that it does, to an extent; however, it would be foolish to assume a direct mapping of one onto the other. Distractor congruence, for example, includes a top-down influence of the category to which distractors and the search target belong; standard models of target-distractor similarity may be unable to account for this added influence.

Another question is of how to adapt models of search to fit the effects underlying their correct predictions. Certainly it could be argued by Ockham's Razor that, given the excellent fit of the models to the overall data, there need not be further investigation of these effects. I would make the counterpoint, however, that this strategy will fail for visual search that falls outside of the norm. Distractor congruence has proven to be a powerful predictor which elegantly explains much of the variance underlying typical search patterns in naturalistic imagery; for this reason, it could likely be implemented in current models of search without difficulty.

When considering other components for theories of visual search such as FIT, GS, and AET, it is perhaps worth considering the size of the effects used by the models.

In Experiment 2, the background accounted for a mere 7% of the variance, despite being an incredibly robust effect. Image class, on the other hand, accounted for well over 30% of the variance, and distractor congruence proved to be a strong predictor of response time on its own. That is not to say that these models should abandon the effects of the background, as the predictor is still robust; rather, a prudent course of action would be to try to integrate both distractor congruence and background complexity into current models of visual search.

In order to do this, the nature of the effect of distractor congruence (and perhaps image class) will need to be studied in more depth. While the results found on these 32 images were incredibly robust, 32 images is a drop in the bucket of the scope of naturalistic visual search. Indeed, perhaps the most prudent suggestion for future research is to do additional studies in naturalistic imagery on distractor congruence. Such research should follow the guiding principle that control should only be maximized insofar as the naturalistic nature of the scene can be maximized. Thus, rather than studying purely naturalistic imagery or purely sanitized lab imagery, I suggest that the most useful research will find its home somewhere in the middle. Perhaps the knowledge gained from such studies of this “middle ground” can yield insights crucial to proper control in purely naturalistic settings.

7 References

Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search: The Frankenbear experiment. *Vision research*, *54*, 20-30.

Bamber, D. (1969). Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception & Psychophysics, 6*(3), 169-174.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*, *14*(2), 143-177.

Biggs, A. T., Cain, M. S., Clark, K., Darling, E. F., & Mitroff, S. R. (2013). Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Visual Cognition*, *21*(3), 330-352.

Bird, R. E., Wallace, T. W., & Yankaskas, B. C. (1992). Analysis of cancers missed at screening mammography. *Radiology*, *184*(3), 613-617.

Boot, W. R., Neider, M. B., & Kramer, A. F. (2009). Training and transfer of training in the search for camouflaged targets. *Attention, Perception, & Psychophysics*, *71*(4), 950-963.

Caird, J. K., Edwards, C. J., Creaser, J. I., & Horrey, W. J. (2005). Older driver failures of attention at intersections: using change blindness methods to assess turn decision accuracy. *Human Factors*, *47*(2), 235-249.

Cathcart, J. M., Doll, T. J., & Schmieder, D. E. (1989). Target detection in urban clutter. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(5), 1242-1250.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive psychology*, *22*(2), 225-271.

Clark, K., Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2012). Overcoming hurdles in translating visual search research between the lab and the field. *The influence of attention, learning, and motivation on visual search* (pp. 147-181). Springer, New York, NY.

Davies, G., & Hine, S. (2007). Change blindness and eyewitness testimony. *The Journal of psychology*, *141*(4), 423-434.

Driver, J., McLeod, P., & Dienes, Z. (1992). Motion coherence and conjunction search: Implications for guided search theory. *Attention, Perception, & Psychophysics*, *51*(1), 79-85.

Donderi, D. C. (2006). Visual complexity: a review. *Psychological bulletin*, *132*(1), 73.

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

Duncan, J., & Humphreys, G. (1992). Beyond the search surface: visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance. 18*(2), 578-588.

Duncan, J. (1995). Target and nontarget grouping in visual search. *Attention, Perception, & Psychophysics*, *57*(1), 117-120.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (No. 3). *University Microfilms.* Teachers College, Columbia University.

Farmer, E. W., & Taylor, R. M. (1980). Visual search through color displays: Effects of target-background similarity and background uniformity. *Attention, Perception, & Psychophysics*, *27*(3), 267-272.

Fryklund, I. (1975). Effects of cued-set spatial arrangement and target-background similarity in the partial-report paradigm. *Attention, Perception, & Psychophysics*, *17*(4), 375-386.

Galpin, A., Underwood, G., & Crundall, D. (2009). Change blindness in driving scenes. *Transportation research part F: traffic psychology and behavior*, *12*(2), 179-185.

Halverson, T., & Hornof, A. J. (2004). Local density guides visual search: Sparse groups are first and faster. *Proceedings of the human factors and ergonomics society annual meeting. 48*(16), 1860-1864.

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of experimental psychology: Human perception and performance*, *25*(1), 210.

Ho, G., Scialfa, C. T., Caird, J. K., & Graw, T. (2001). Visual search for traffic signs: The effects of clutter, luminance, and aging. *Human Factors*, *43*(2), 194-207.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8*(4), 761-768.

Ishihara, I. (1980). Ishihara's Tests for Color-Blindness: Concise Edition. Tokyo: Kanehara & Co., LTD.

Judd, T., Durand, F., & Torralba, A. (2012). A benchmark of computational models of saliency to predict human fixations.

Krupinski, E. A. (2005). Visual search of mammographic images: Influence of lesion subtlety. *Academic radiology*, *12*(8), 965-969.

Kundel, H. L., & La Follette Jr, P. S. (1972). Visual search patterns and experience with radiological images. *Radiology*, *103*(3), 523-528.

Lavie, N., & Cox, S. (1997). On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. *Psychological Science*, *8*(5), 395-396.

Mack, A., & Rock, I. (1998). *Inattentional blindness.* 33. Cambridge, MA: MIT press.

Menneer, T., Cave, K. R., & Donnelly, N. (2009). The cost of search for multiple targets: effects of practice and target similarity. *Journal of Experimental Psychology: Applied*, *15*(2), 125.

Mogelmose, A., Trivedi, M. M., & Moeslund, T. B. (2012). Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. *IEEE Transactions on Intelligent Transportation Systems*, *13*(4), 1484-1497.

Most, S. B., & Astur, R. S. (2007). Feature-based attentional set as a cause of traffic accidents. *Visual Cognition*, *15*(2), 125-132.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of neuroscience methods*, *222*, 250-259.

Neider, M. B., & Zelinsky, G. J. (2006a). Scene context guides eye movements during visual search. *Vision research*, *46*(5), 614-621.

Neider, M. B., & Zelinsky, G. J. (2006b). Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision research, 46*(14), 2217-2235.

Neider, M. B., Boot, W. R., & Kramer, A. F. (2010). Visual search for real world targets under conditions of high target–background similarity: Exploring training and transfer in younger and older adults. *Acta Psychologica*, *134*(1), 29-39.

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, *11*(14), 7-7.

Nelson, K. J., Laney, C., Fowler, N. B., Knowles, E. D., Davis, D., & Loftus, E. F. (2011). Change blindness can cause mistaken eyewitness identification. *Legal and criminological psychology*, *16*(1), 62-74.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003, September). Top-down control of visual attention in object detection. *Proceedings 2003 International Conference on Image Processing*. *3*(1), 250-253. IEEE.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of vision*, *7*(2), 17-17.

Raff, M. S. (1950). A volume warrant for urban stop signs. *Traffic Quarterly, 4*(1), 48-58. Eno Transportation Foundation.

Rubin, E. (2001). Figure and Ground. *Visual Perception*. 225-229. Philadelphia, Psychology Press.

Rumar, K. (1990). The basic driver error: late detection. *Ergonomics*, *33*(10-11), 1281-1290.

Sareen, P., Ehinger, K. A., & Wolfe, J. M. (2016). CB Database: A change blindness database for objects in natural indoor scenes. *Behavior research methods*, *48*(4), 1343-1348.

Schmieder, D. E., & Weathersby, M. R. (1983). Detection performance in clutter with variable resolution. *IEEE Transactions on Aerospace and Electronic Systems*, (4), 622-630.

Scialfa, C., Esau, S., & Joffe, K. (1998). Age, target-distractor similarity, and visual search. *Experimental Aging Research*, *24*(4), 337-358.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 194.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological review*, *95*(1), 15.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of experimental psychology: human perception and performance*, *16*(3), 459.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 652.

Treisman, A. (1992). Spreading suppression or feature integration? A reply to Duncan and Humphreys (1992). *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 589-593.

Virkler, K., & Lednev, I. K. (2009). Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International*, *188*(1), 1-17.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 419.

Wolfe, J. M. (1994a). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Wolfe, J. M. (1994b). Visual search in continuous, naturalistic stimuli. *Vision research*, *34*(9), 1187-1195.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision research*, *42*(28), 2985-3004.

Wolfe, J. M. (2007). Guided search 4.0. *Integrated models of cognitive systems*, 99-119.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, *136*(4), 623.

Wolfe, J. M. (2010). Visual search. *Current biology*, *20*(8), R346-R349.

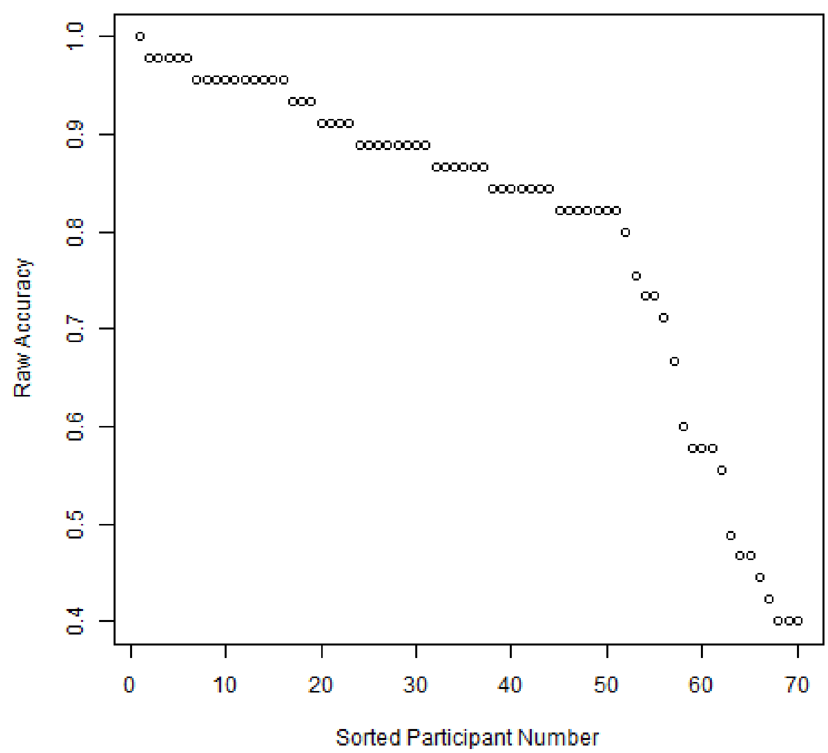
Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121-124.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of vision*, *13*(3), 33-33.

8 Appendices

8.1 Appendix I – Data Trimming, Experiment 1

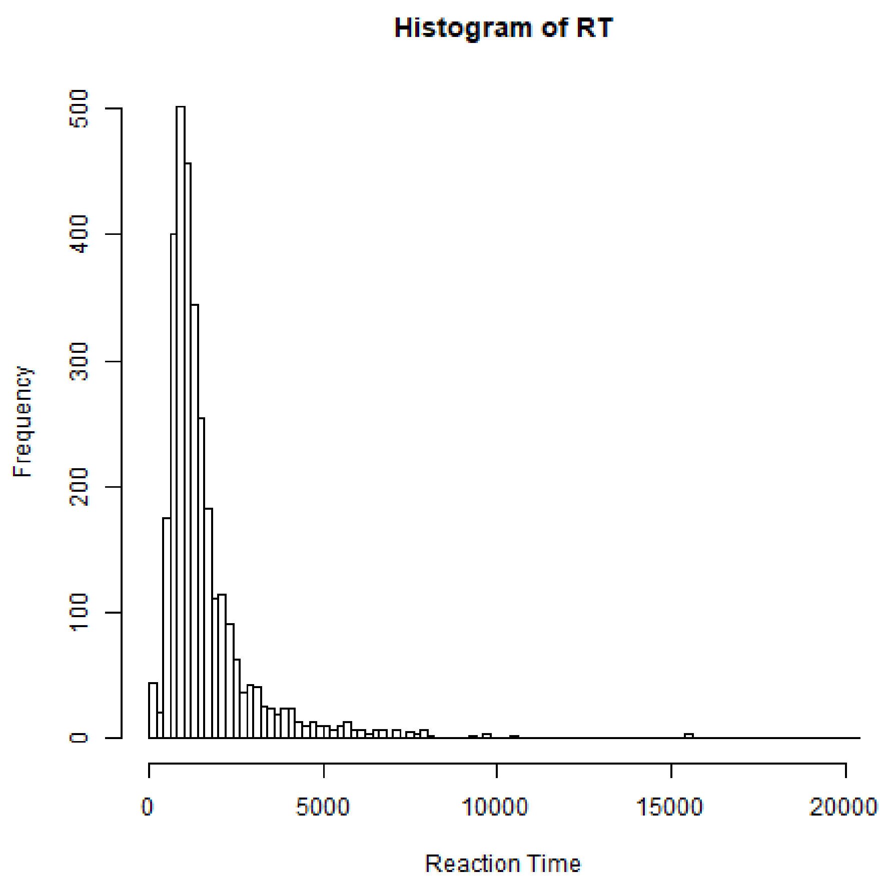
With such a wealth of data, the task of data trimming must be done with great care. Data were trimmed based on two metrics – first on a participants' individual accuracy (allowing the removal of those who didn't do the task correctly) and then on RT on a trial-by-trial basis.



Graph 1: Raw Accuracy by Participant

In Graph 1, a distinct elbow can be seen after participant 52, breaking away from the pattern of plateaus and falling sharply. For this reason, participants 53-70 were omitted from analysis.

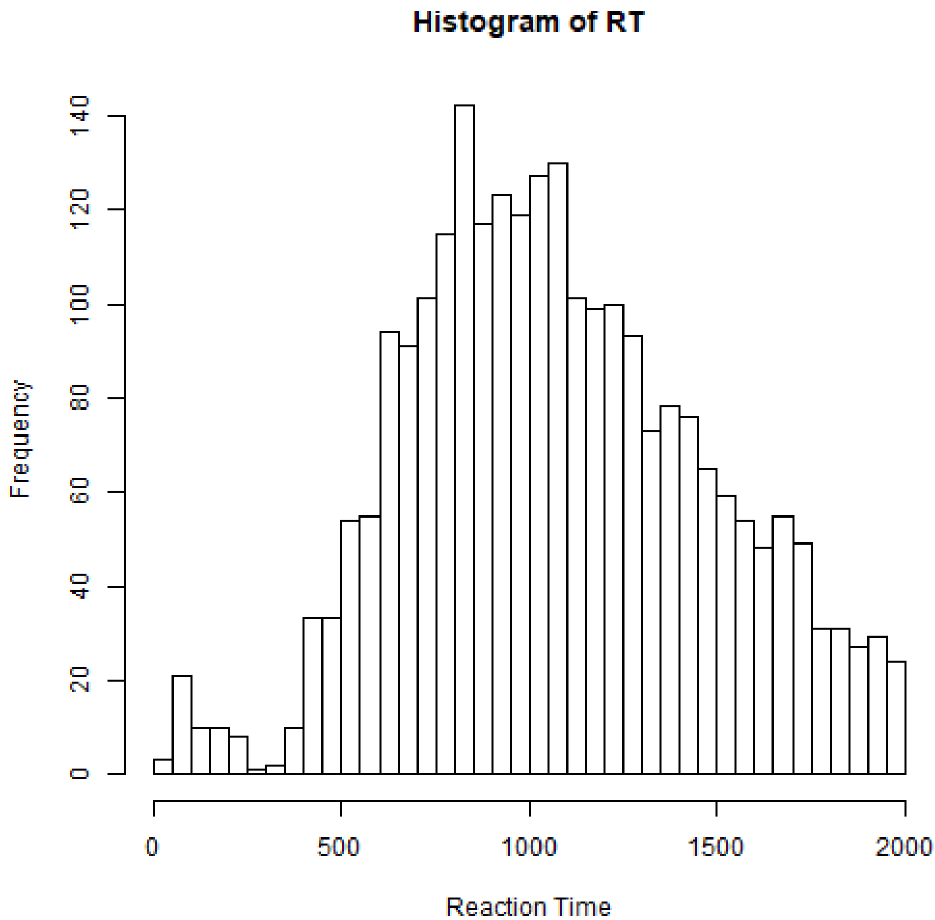
Appendix I continued



Graph 2: Raw Reaction Time

With inaccurate participants removed, the next step was to remove unreasonable response times from the mix. Two trims needed to be made: one for false starts, and one for unreasonably long trials.

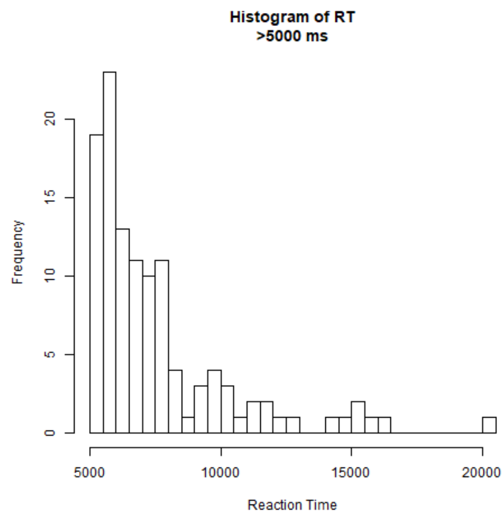
Appendix I continued



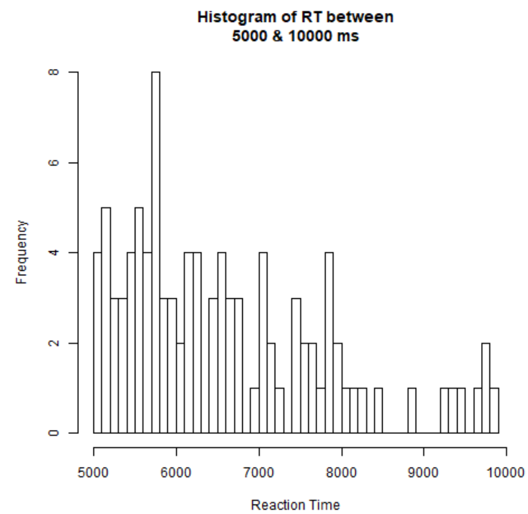
Graph 3: Lowest Raw Reaction Times

As can be seen in Graph 3, there is a small conglomerate of responses that are much faster than the rest. These responses were attributed to false starts, and were removed accordingly. A threshold of 350 ms was established as the cut-off point.

Appendix I continued



Graph 4: Highest Reaction Times



Graph 5: Highest RT's Under 10,000ms

In Graph 4, there appeared to be a reasonable cut-off point just before 10,000 ms, but it wasn't clear where specifically it was; while cutting the data after about 14,000 ms would not be unreasonable, there appeared to be some noise around the 10,000 ms mark. Thus, a closer look was warranted. Graph 5 shows a zoomed in look at the segment in question, and based on this it seemed reasonable to cut the data after 8,500ms.

This completes Appendix I – Data Trimming, Experiment 1

8.2 Appendix II – Imagery & Search Targets, Experiment 1

8.2.1 Imagery

Example picture of each background. For each background, fifteen trials were conducted, separated into five distinct trials per target. Participants were shown 1 of the colored targets to study, and then were shown the large image with up to 2 of the targets appearing in color.

Appendix II continued

8.2.2 Complete List of Search Targets

Image A Search Targets

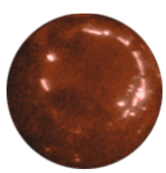
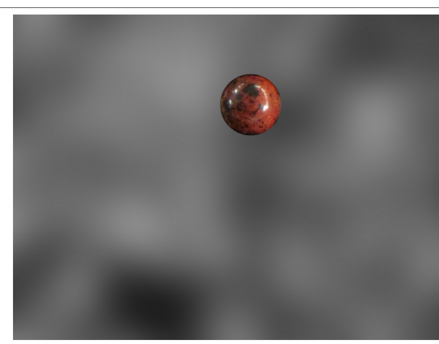
Image B Search Targets

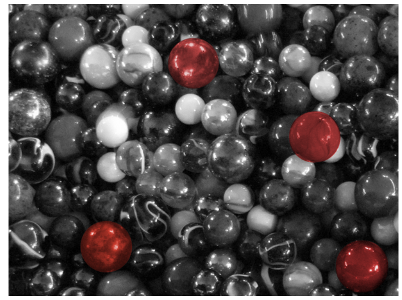
Image C Search Targets

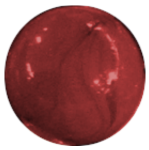
This completes Appendix II – Imagery & Search Targets, Experiment 1

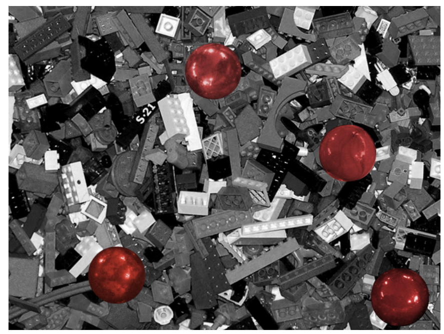
8.3 Appendix III – Imagery & Search Targets, Experiment 2

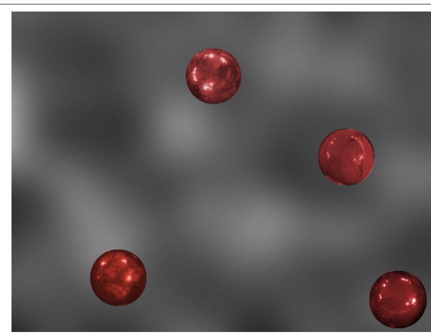
Each participant saw each target exactly once, with each target pair appearing on three potential backgrounds. Participants would see two backgrounds from each set of three during the experiment – one for each target. Participants were never shown the same background twice. See Appendix IV for information on the complete set of imagery.

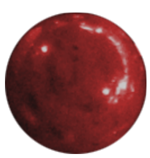
Target to Search For	**Imagery to Search Within**
	
	
	

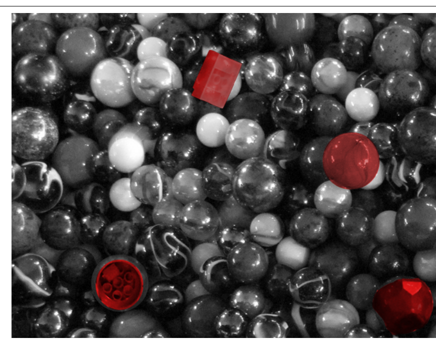


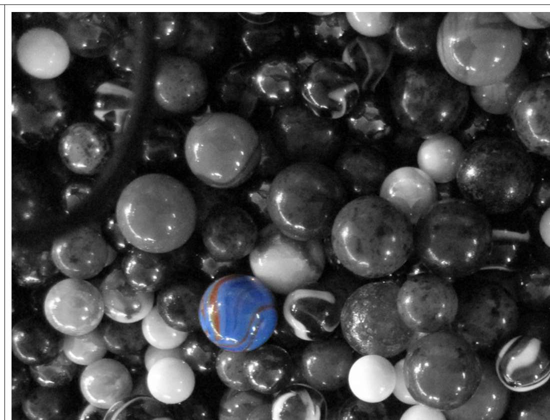


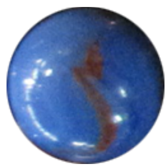


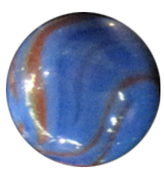


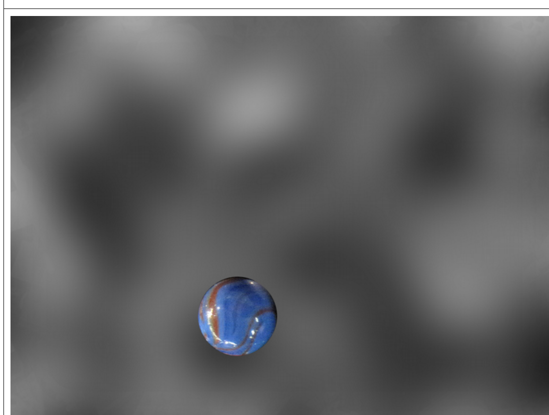


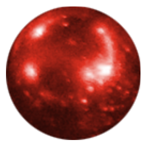


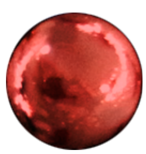


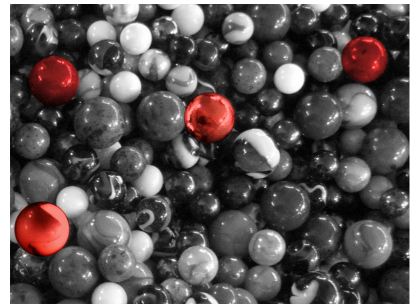


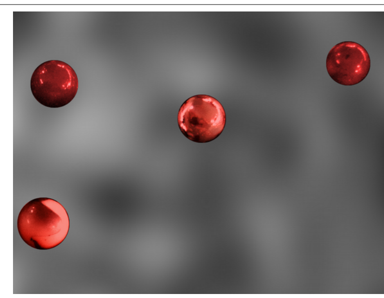


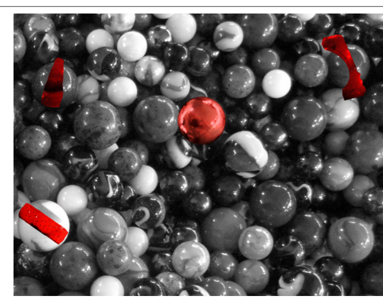












8.4 Appendix IV – Copyright and Full Image Set

The set of imagery used in this experiment, as well as specific copyright information about each image, is too large and complex to reasonably include in this document. The full image set and copyright information are hosted online at Zenodo.org, and can be accessed by anybody who creates a free account.

Images are hosted at the following link: https://zenodo.org/record/1219145

DOI: 10.5281/zenodo.1219145

Please note that images in this image set fall under several different licensing requirements. Information about these requirements can be found in the 'READ ME.doc' and 'Image Copyright Information.xls' files included in the image set. Use of any part of this image set constitutes acknowledgement of the specific license requirements of each image and agreement to abide by these requirements.