Summer 2016

Methods Of Measuring Visual Scanning Of Upright And Inverted Ecological Images

Benjamin Lee Graves

METHODS OF MEASURING VISUAL SCANNING OF UPRIGHT AND INVERTED ECOLOGICAL IMAGES

A Masters Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science, Psychology

By

Benjamin L. Graves

July 2016

METHODS OF MEASURING VISUAL SCANNING OF UPRIGHT AND INVERTED ECOLOGICAL IMAGES

Psychology

Missouri State University, July 2016

Master of Science

Benjamin L. Graves

ABSTRACT

Facial recognition has been long held as a special perceptual process at which humans excel, and is primarily a function of perceptual experience. However, there are experimental manipulations that impede this perceptual process and make it more difficult for humans to recognize the face (i.e. only presenting half a face or inverting the face). In the case of inversion, it is though that the inverted face interrupts a person's ability to process the face holistically and forces a change to featural processing. The purpose of this experiment was to examine if inversion of ecologically valid images would also impact recognition memory. In this study, individual differences in adult participant's natural propensity to scan, recognition memory response latency, and recall memory for upright and inverted urban and office scenes was investigated. Overall, using a 2 (Group: Upright versus Inverted) x 3 (Trail Block) design, it was found that visual scanning rate tended to be faster for upright versus inverted images, recognition memory response latencies were significantly slower for inverted images, and rates of fixation tended to decrease across trial blocks. However, differences in fixation rates arose when assessing natural propensities to scan and during the item recall task.

KEYWORDS: visual scanning, ecological stimuli, recognition memory, scene perception, individual differences

This abstract is approved as to form and content

____

D. Wayne Mitchell, PhD
Chairperson, Advisory Committee
Missouri State University

METHODS OF MEASURING VISUAL SCANNING OF UPRIGHT AND INVERTED ECOLOGICAL IMAGES

By

Benjamin L. Graves

A Masters Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Science, Psychology

July 2016

Approved:

D. Wayne Mitchell, PhD

Erin M. Buchanan, PhD

Russell N. Carney, PhD

Julie Masterson, PhD: Dean, Graduate College

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis committee, Dr. Wayne Mitchell, Dr. Erin Buchanan, and Dr. Russell Carney, for giving me their invaluable time and advice throughout the thesis process. They each played their own integral role in the development of not only this project but also my experience at Missouri State, which would not have been possible without them. Additionally, I am grateful for the MSU Graduate College for providing funding for supplies used in the study.

I would also like to thank the awesome faculty and staff in the Psychology Department that contributed to my academic growth the past two years and helped keep me afloat while gaining new skills. Finally, thanks go out to my family and past educators for providing support and encouragement that has led to my academic success.

TABLE OF CONTENTS

Introduction 1
Concepts of Visual Attention 1
Attention to Faces and the Facial Inversion Effect 4
Attention to Scenes 6
Measures of Visual Perception 9
Purpose of Study 10

Method 11
Participants 11
Stimuli 11
Apparatus and Measurements 11
Procedure 13

Results 16
Phase 1 Analyses 16
Phase 2 Analyses 19
Phase 3 Analyses 23

Discussion 27
Phase 1 Discussion 27
Phase 2 Discussion 29
Phase 3 Discussion 30
Conclusions 31

References 33

Appendix. Consent Form and Demographic Information Sheet 37

LIST OF TABLES

Table 1. Scan path similarity summary...17

Table 2. *Phase 1* fixation rate summary 18

Table 3. *Phase 2* response latency summary 22

Table 4. *Phase 2* fixation rate summary 23

Table 5. *Phase 2* length of first fixation summary...24

LIST OF FIGURES

Figure 1. String edit similarity 12

Figure 2. *Phase 1* fixation rate interaction 18

Figure 3. *Phase 1* fixation rate scatter plots 20

Figure 4. Scan path similarity scatter plots 21

Figure 5. Response latencies across trial blocks 22

Figure 6. Fixation rate change across trial blocks 24

Figure 7. *Phase 3* group fixation rate 25

Figure 8. *Phase 3* scatter plots 26

INTRODUCTION

A vast amount of research has been conducted to assess how humans perceive and recognize faces, and many researchers argue that there are special subcortical, perceptual, and/or cognitive processes that account for face recognition (Johnson, Senju, & Tomalski, 2015; Richler & Gauthier, 2014; Nakabayashi & Liu, 2014). However, very few studies have been conducted on ecological scenes using similar paradigms to face recognition memory studies. One such paradigm is used to study the facial inversion effect (FIE). These studies often rely on very fast (e.g., 250-500 milliseconds) stimulus presentations and use response latency (RL) as a dependent measure; however, advances in eye-tracking technologies provide alternative ways of measuring individual differences in perceptual processing. The FIE paradigms, along with modern eye-tracking devices and derived measures, can be used to ascertain if similar perceptual effects arise when using ecological images, such as urban or natural scenes. The subsequent introduction will (1) provide a brief overview of concepts of visual attention, (2) review theories of attention for facial and scene perception, (3) discuss measurements that can be derived from eye-tracking devices and their limitations, and (4) provide a rationale for the current study.

Concepts of Visual Attention

Attention encompasses a variety of complex mental processes that afford humans the ability to allocate and direct processing resources to stimuli in their environment. The development of attention is thought to occur as a two-stage process as defined by James

(1890). It starts as an involuntary process, passively scanning the environment under control of novel and/or salient stimuli, and then becomes more selective with the person actively directing their attention toward the stimuli of interest or importance. In the first stage, early saccadic eye movements are more passive, while post saccadic eye movements are more active, suggesting a shift in visual attention (Mathôt & Theeuwes, 2011). Feature integration theory, as suggested by Treisman and Gelade (1980), offers an explanation to this process. It posits that features in the environment are perceived first in a very quick and efficient manner, which is often referred to as holistic processing. The features are then turned into objects, which are identified as the allocation of feature detection starts to narrow, and finally, attention is then directed to the stimulus or stimulus attributes of importance. With this process, objects that are distinct from other objects in the environment can be perceived readily during the initial stage, while objects that share conjunctive (i.e. matched) features with other objects or are missing a feature must be processed serially. It is assumed here, that stimulus orienting and the initial visual processing of a stimulus begin with preattentive processes – the term first coined by Neisser (1976). This orienting can be interpreted as stimulus control, an aspect of J. J. Gibson (1966), and can be viewed as a component of Solokov's (1963) orienting reflex.

Preattentive processing is guided by stimulus saliency, contrast, novelty, motion, and/or size. Active, voluntary featural processing follows the preattentive processing stage. Preattentive processing allows for quick acclimation of information from the environment. It is a very fast allocation of attention to properties of the environment, almost in a parallel processing method. The individual's gaze is then directed to areas of novelty, high saliency, or certain configurations that are of important functionally. With this, initial saccadic eye movements, that represent preattentive processing, have very quick scanning rates with short fixation durations. As the individual habituates to, or becomes familiar with, the environment, scanning rate decreases and fixation durations increase, suggesting an allocation of attentional resources to the most important stimulus component at that moment.

Since only basic features are being attended to, preattentive processing is thought to be bottom-up. Viewed features activate feature detectors in the brain that analyze information, and these detectors appear to increase in the amount of information they are able to process with age (Treisman, Vieira, & Hayes, 1992). In primates, these feature detectors seem to have their own neural networks, with most occurring in the striate cortex (DeYoe & Van Essen, 1988). The preattentive process is illustrated in the way infants develop adult-like scanning skills. Initial infant scanning, at about 6 weeks of age, is directed by the saliency of features in the stimulus, but as they get older, at around 13 weeks, scanning becomes more selective and shows more control over targets (Bronson, 1994). Another conclusion that can be drawn from this research is that preattentive abilities improve and develop over time, and that features can be retained in memory for future use. However, preattentive processes should not be confused with automatic processes. Automatic processes take time and practice to develop, while preattentive processes are controlled by innate mechanisms and are acquired early in life (Treisman et al., 1992).

Preattentive processing guides initially a person’s perceptions when they are orienting to a new stimulus and then helps to direct later selective attention to areas of importance and allows for rapid, holistic processing and easier recognition of the stimulus. For faces, importance is typically the areas around the eyes, while for scenes it is areas of high context or salience. However, unexpected changes in the stimulus can impede these processes and result in the perception and recognition of the stimulus to be impaired.

Attention to Faces and the Facial Inversion Effect

Evolutionarily, humans have become very well adapted to the perception and recognition of faces. The attention and perception of faces is known to be a special process that occurs neurologically as an innate ability and can be readily seen in infancy as two subcortical processes (Morton & Johnson, 1991). Morton and Johnson (1991) argue that facial perception is an innate system that provides humans with a predisposition to orient towards, detect faces, and fosters facial recognition. While these processes are easier to see and separate out in the human infant, researchers have shown that adults also show a preference to orient toward faces in a scene and facial recognition skills can still develop over time (Johnson, Senju, & Tomalski, 2015). One possible explanation for face recognition among perception researchers is holistic processing. The holistic representation of faces appears evident during the first year of life through adulthood (see Richler & Gauthier, 2014, for a review). However, using the term "holistic" to describe processing methods comes with its own issues. Researchers have provided varying definitions and ways to measure holistic processing, especially for faces, and even though attempts have been made to mathematically operationalize the process, theories are having a difficult time converging. However, holistic processing is represented by two different research camps, those supporting the configural-relational processing model versus those supporting the part-based model (see Richler, Palmeri, & Gauthier, 2012, for more). In the configural-relational processing model, researchers have proposed that perceptual processing relies primarily upon the spacing and position of the face features, whereas in the part-based model, perceptual processing is led primarily by parallel processing all parts in a holistic fashion (a "whole is greater than the sum of the parts" definition).

There are many factors that influence facial recognition, some of which also impact the performance of facial recognition algorithms (Lui, Bolme, Draper, Beveridge, Givens, & Phillips, 2009). Manipulations such as how the face is posed, cognitive load, race, attention duration, and transformations can promote or inhibit a person's ability to recognize a face (Shapiro & Penrod, 1986; Reynolds & Pezdek, 1992). One transformation that occurs often in research is facial inversion. In these studies, participants are shown inverted faces to encode and then asked if they remember the face at a later recognition task. These results are compared to a group that received a similar condition, but with upright faces. It is theorized that when face images are inverted, it is difficult to process the spatial relationships between features and causes face recognition to become impaired. This result has been coined the facial inversion effect (FIE).

The consistent finding of the FIE is that facial recognition for upright facial images are more accurate and faster than for recognition of inverted facial images (Yin, 1969, 1970). The basis of the holistic hypothesis is that faces are processed as a whole object and it is difficult to break the face down into its individual features. When faces are inverted, the holistic process is disturbed and recognition of the face is impaired; however, featural recognition remains unaffected (Rakover, 2013). This effect is lessened when only features are used, but when the feature being focused is masked and the rest of the face can be seen in the periphery, the FIE is increased (Van Belle, De Graef, Verfaillie, Rossion, & Lefèvre, 2010).

Cashon and Cohen (2003) provide an interesting finding in the development of the FIE in infants. In their research, infants begin processing upright and inverted faces by using features and by the time they reach an age of four months they are processing both holistically. However, between four and six months of age infants revert back to processing the faces by features which suggests the system is being overloaded and must return to a lower processing level possibly due to a combination of factors such as an increase in visual acuity and/or an increase in knowledge of social significance. By 7 months of age, the infants regain the ability to processes upright faces holistically, but not inverted faces. The authors suggest that this differentiation could be due a change in how the environment is now perceived. By this age, infants' ability to consistently sit upright and continuously see faces in an upright orientation (i.e. eyes above the nose and mouth). It could be assumed that there is a similar impairment in recognition and change in attention when viewing images or scenes of everyday environments; however, this idea has yet to be tested.

Attention to Scenes

Although the way faces are perceived is distinctive, there is most likely similarity and overlap of how scenes are perceived. When viewing an upright image, an individual will tend to process the scene in a holistic manner. Using holistic processing allows for the rapid processing of a scene and explains how people are able to get the gist of a scene (a basic understanding of what the scene is) with very quick image presentations or blurred images (Potter & Levy, 1969; Oliva & Torralba, 2006). It is hypothesized that when an image is inverted, however, individuals will shift from the holistic processing to a more featural based processing method. This switch should demonstrate a similar effect to the FIE whereby the inverted images cause a disruption in holistic processing abilities.

Some researchers have made efforts to predict what areas of an image an individual will fixate on or attend to based on saliency, context, and other internal top-down factors. Saliency models suggest that gaze is directed to the most salient areas of the image and are driven by basic features. These are often areas that provide the most contrast between features in the image. The models are rooted in feature-integration theory, and are based on the idea that attention starts with the low-level features and then combines slowly the features into object. This model suggests a bottom-up approach to image attention in which the viewer processes serially areas of the image starting with the most salient and moving down. However, there is an issue that comes with this bottom-up approach. Saliency models seldom take into account any holistic information from the scene, leaving out the learning aspect of the attentive process that turns the objects into features for quick identification of the scene. They treat anything that is not the target of search as noise that distracts the perceiver and complicates the search process (Torralba, 2003). These models do predict recognition memory better than random models, but only slightly (Foulsham & Underwood, 2008).

Contextual models of attention use the association of objects in the environment to predict where an individual will look. Since objects do not occur in isolation and associations can be formed between objects, people can use contextual cues to quickly process scenes and guide attention to where it is needed most (Oliva & Torralba, 2007). Statistical regularities in scenes direct attention to areas in the scene where a certain object is likely to be found. They can also provide insight into how large or small the object should be. Contextual models take into account three aspects to determine where attention will be directed: saliency, target-driven control, and contextual priors. Saliency is similar to the models mentioned previously and control for the bottom-up factors of attention. Target-driven control is the top-down knowledge of where an object would normally appear in a scene. So, areas that are more likely to have the object will receive more attention than areas that are unlikely to contain it. Prior contextual experience provides information about the object, such as what it is, where it should be located, and its scale (Torralba, 2003). Taking all of these factors into consideration, and using a holistic model instead of solely featural, provides a much better model to predict where vision will be directed in a scene.

When images are inverted, attention should change from being more contextually based (holistic) to more saliency based (featural). Findings by Harding and Bloj (2010) suggest that scan paths are slightly disrupted when images are inverted or the luminance of the image is changed. Their research suggests that the individuals tend to scan the same image in a similar pattern during multiple presentations even when there are certain holistic changes. Scan paths will be discussed in greater detail below. So, it seems that image inversion does disrupt the holistic process, but measures have to be developed and refined to be indicative of this phenomenon.

Measures of Visual Perception

The majority of published works assessing aspects of visual attention and perception have been based upon tasks that employ stimulus manipulations using RL as the primary dependent variable. The differences in RL (faster or slower as function of stimulus manipulation) were used to support the type of perceptual processes employed in face recognition. Where an individual actually directed their visual attention was not measured, just assumed. With development of modern eye-tracking devices and software, researchers have been able to better test hypotheses such as the ones presented above. However, the visual scanning measures employed typically are those pre-defined by the eye tracking company's software and/or are selective based upon a researcher's preference. Hence, concordance of visual scanning measurement among researchers is lacking; and too, such differences in visual scanning measurement can result in differences in study results (Phillips & Edelman, 2008, Foulsham & Underwood, 2008).

Visual scanning software provides information on locations of fixations via coordinates, their durations, and other technical aspects, but often these can be difficult to use in analysis. Using areas of interest (AOIs) allows for more useful information to be obtained such as the number of fixations in a certain area, the latency to respond to a given location, and how many times they leave and return to the area. Such visual scanning data provides little information about the actual perceptual process(es) and changes in the perceptional process(es). However, other measures can be derived from the stock measures such as scan paths analyses, latency to first fixation, and scanning rates. Scan path analyses allow the researcher to find similarities in the scanning patterns of the participants. Latency to the first fixation is the time it takes for the participant to move their gaze from the starting point to the first fixation point and has been suggested to be an indicator of how fast RL will be on a recognition memory task. Scanning or fixation rates illustrate differences in the speed that a person is scanning an image. It is calculated by the number of fixations per unit of time, usually seconds or milliseconds. Changes in fixation rate could suggest differences in how a person is attending to an image. A description of these derived measures used in this study will be presented in the method section.

Purpose of the Study

The purpose of this study is fourfold: (1) to test the inversion effect with ecologically valid, real world, stimuli; (2) to assess the stability of various derived measures of visual scanning on upright versus inverted stimuli; to date stability of visual scanning has yet to be explored; (3) to assess the relationship between visual scanning and subsequent recognition memory on upright versus inverted stimuli; and (4) to assess the relationship between visual scanning and recall memory for upright and inverted stimuli. It is hypothesized that scan paths and fixation rates should be more consistent for upright images compared to inverted images and fixation rates should be faster for upright images than inverted. Recognition memory RLs will decrease across trials, but will be slower for inverted images compared to that of upright images. The number of items recalled should be better for the inverted images than the upright images.

METHOD

Participants

Prior approval for this project was obtained from the Missouri State University IRB (April 26, 2016; approval # 16-0416). Fifty-nine undergraduate participants recruited through the SONA system and from other psychology classes with professors' permission were used in the study. Three participants were removed; 1 due to experimenter error, 1 due to equipment malfunction, and 1 due to instructional misunderstanding, leaving a total $N = 56$ participants.

Stimuli

Stimuli consisted of 11 urban images obtained from the LabelMe image database (Russell, Torralba, Murphy, & Freeman, 2008) and one modern office image obtained from a Google search. The images were presented, in a landscape orientation, on a 60 cm computer monitor in full screen mode. However, the presentation size was limited by the vertical or horizontal bounds of the monitor.

Apparatus and Measurements

Eye movements and fixations were recorded by a GazePoint GP3 Eye Tracker mounted to the bottom of the monitor and controlled by GazePoint Analysis and Control software. The visual scanning dependent measures consisted of three derived measures: fixation rate, length of the first fixation (LFF), and a scan path score. Fixation rate was calculated using the number of fixations divided by the total amount of time the stimulus was presented on screen in milliseconds. The number of fixations was determined by using an Area of Interest (AOI) over the entire stimulus and exporting the fixation count data from the GazePoint software program. Length of the first fixation, in milliseconds, was obtained via the GazePoint software, with the constraint that the fixation length was greater than 180 milliseconds. The scan path analysis was conducted using a string-edit similarity method (see Figure 1). For the analysis only, a 5x5 grid of equally sized rectangles was superimposed on the stimulus of interest. Each rectangle of the grid is assigned a letter (A-Y) which was used to create a string of letters denoting the order of areas that the participant fixated on. A letter string was created for each presentation of the stimulus. The letter strings were then compared for similarity by taking the number of edits it took to convert the first string into the second string. The edits could only consist of letter insertions, deletions, or substitutions. The number of edits were then divided by the total number of letters in the string and then subtracted from 1 to get the

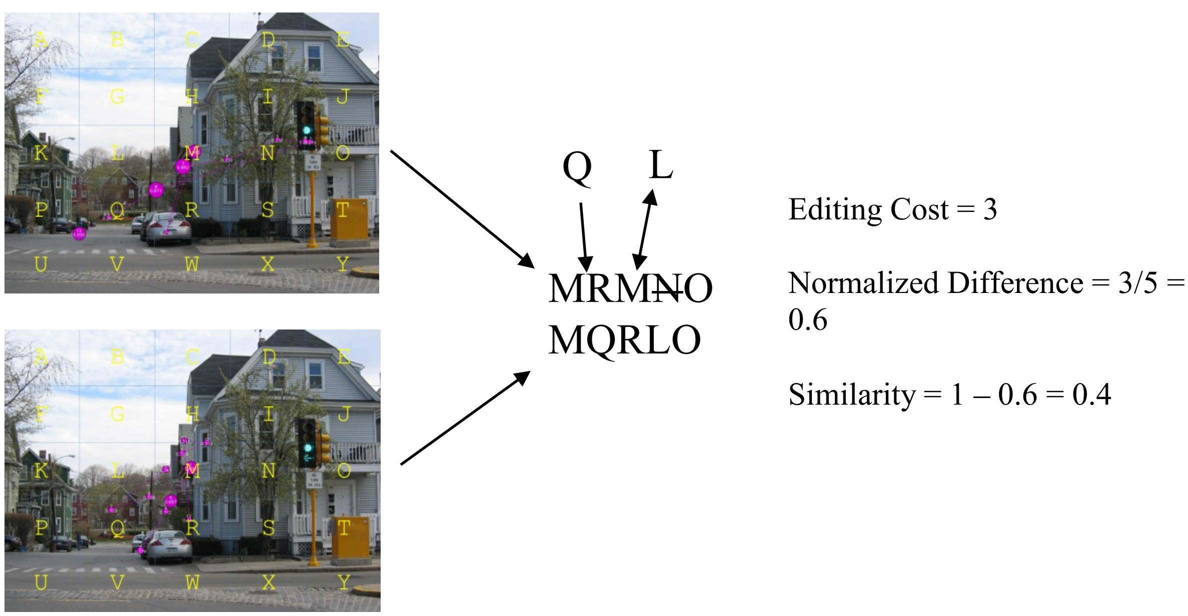

Figure 1. String edit similarity. This figure shows an example of the string edit similarity calculations. To turn string one into string two, it costs three edits. This is then turned into a normalized difference and subtracted from one to get the similarity.

similarity (see Brandt & Stark, 1997, for a more detailed explanation). String edit similarities were calculated in *R* using the *RecordLinkage* package (R Core Team, 2016; Borg & Sariyar, 2016).

Procedure

Participants were assigned randomly to one of the four groups; which viewed inverted or upright images and received a forward or backward stimulus presentation order. Hence, a 2 (Group: Inverted vs. Upright) X 2 (Presentation Order: Forward vs. Backward) between-subjects factorial design was used. Although not mentioned here, the research design employed a repeated measures factor, to be discussed later in the procedure section.

Before testing began, each participant was given a brief overview of the purpose of study. Each participant was then given the Informed Consent Form and a Demographic Information Sheet (see Appendix A). Once consent was obtained, the participant was led into the testing room and seated approximately 61 cm in front of the image display monitor (GazePoint suggested guidelines). The participant was instructed to find a relaxed position and avoid excessive movement. Eye tracking calibration procedures were then conducted using GazePoint's 5-point calibration process. Calibration is necessary to guarantee accurate tracking of the participant's eyes in concordance with the image parameters so to obtain reliable visual scanning data. Once calibration was successful, the testing began. There were three test phases in this experiment. Each phase is discussed in turn in the following sections.

Phase 1 was designed to evoke and capture the participant's natural scanning propensities. Each participant was shown the same urban scene image, either upright or inverted depending upon group assignment, three times. The image was presented at a resolution of 1600 x 1200 pixels. The participants were instructed to just look at the image on the screen and that they were free to look wherever they wanted. The duration of each image presentation was 5 seconds with an inter-stimulus-interval (ISI) of 1 second. During the ISI a black screen was presented with a small white cross (+) centered on the screen. Participants were instructed to continue to look at the white cross during the ISI to maintain the calibration data norms.

Phase 2 consisted of five visual recognition memory problems using a delayed match-to-sample task. Each recognition memory problem involved the presentation of a novel sample image for 3 seconds, followed by a 5 second delay whereby a black screen was displayed, and then a recognition memory test. The images were presented at a resolution of 2560 x 1920 pixels. During the recognition memory test, the sample image and a new image (of equal complexity, content, and composition as the sample stimulus) were presented unilaterally for 5 seconds, during which the participant is instructed to respond verbally which image was the sample. A numeral 1 (40 font) is visible on the left side of the monitor and a numeral 2 (40 font) was visible on the right side of the monitor. Participants were instructed to respond verbally by saying a "1" or "2" depending upon which image they thought matched the sample. The participant's RL was recorded. The recognition memory problems were followed by a 5 second inter-trial-interval (ITI). Each of the five recognition memory problems were presented twice, counter-balancing the left-right image placement during the recognition test; creating a total of 10 recognition memory trials. The distractors and sample pairings were kept constant through the trials. The 10 recognition memory problems were ordered randomly with the constraint that no recognition memory problem was repeated consecutively. Two test orders are created, a forward and a backward. Depending upon group assignment, each participant was presented inverted or upright recognition memory problems and was presented the recognition memory problems in a forward or backward test order. During the recognition memory delay and ITIs, a black screen was presented with a small white cross (+) centered on the screen. Participants were instructed to continue to look at white cross during the delay and ITI so to maintain the calibration data norms.

Phase 3 served as a manipulation check to assess extent that the participants were encoding the image features. A novel image of an office was presented either upright or inverted, depending upon group assignment, for 3 seconds. The image was presented at a resolution of 2560 x 1920. Immediately after image offset, the participant was asked to name as many items in the image as they could recall seeing during the image presentation. The number of correct items recalled as well as the number of incorrect or mistaken items recalled were recorded.

RESULTS

The primary analysis of *Phase 1* was used to assess individual differences in participants' natural propensity to scan visually inverted and upright images. The goal was to explore consistencies in individuals' scan paths and the differences and trends in fixation rate when viewing upright or inverted images. The analyses of *Phase 2* were used to (1) assess the differences in visual scanning and recognition memory RL as a function of image type (inverted or upright), and (2) to assess the relationship between visual scanning of the sample image and recognition memory; with RL, fixation rate and LFF were used as the primary dependent variables. Finally, *Phase 3* not only served as a manipulation check, but the data allowed for an assessment of the relationship between visual scanning and image items recalled. All analyses were conducted using *R* in conjunction with the *ez*, *reshape*, and *ggplot2* packages (Wickham, 2007, 2009; Lawrence, 2015; R Core Team, 2016)

Phase 1 Analyses

Scan path analysis. Scan path score similarities were calculated, using the string edit method presented previously, between the first and second image presentations and the second and third presentations. Similarities were not assessed between presentations one and three in an attempt to reduce artifacts in the data that could be due to the second stimulus presentation. A 2 (Group: Upright vs. Inverted) x 2 (Stimulus Pair) mixed factorial ANOVA, with repeated measures on the last factor, was conducted using the similarity scores (N = 52). The results did not yield any significant differences in the data, suggesting that scan paths were consistently similar between the upright and inverted groups and also between the stimulus pairs (Table 1).

Fixation rate. Fixation rates were also calculated using the method as defined previously (number of fixations divided by the total amount of time the stimulus is presented in milliseconds) for each image presentation. Participants with incomplete data due to GazePoint malfunction were not included in the analyses ($N = 52$). A 2 (Group) x 3 (Stimulus Presentation) mixed factorial ANOVA, with repeated measures on the last factor, was analyzed on the fixation rates. The data are displayed in Figure 2. The ANOVA did not yield any significant main effects, but did show a significant interaction, $F(2, 50) = 3.11$, $p = .049$, $\eta^2 = 0.03$, between the independent variables (Table 2). *Post hoc* pairwise comparisons, with a Bonferroni correction, were conducted between the groups for each stimulus presentation. A significant difference was found for the second presentation, $p = .025$, $d = 0.64$, between the upright ($M = 0.0027$, $SD = 0.00028$) and inverted ($M = 0.0024$, $SD = 0.00056$) groups, but no significance was found for the first, $p = .54$, $d = 0.17$, or third presentation, $p = .20$, $d = 0.36$, between the upright ($M = 0.0025$, $SD = 0.00042$; $M = 0.0026$, $SD = 0.00041$) and inverted groups ($M = 0.0026$, $SD = 0.00048$; $M = 0.0025$, $SD = 0.00044$). As can be seen in Figure 2, participants' visual

Table 1. Scan path similarity summary.

Effect	df_n	df_d	F	p	η^2
Group	1	50	1.48	.230	0.02
Stimulus Pair	1	50	< 1	.335	0.01
Group: Stimulus Pair	1	50	1.13	.293	0.01

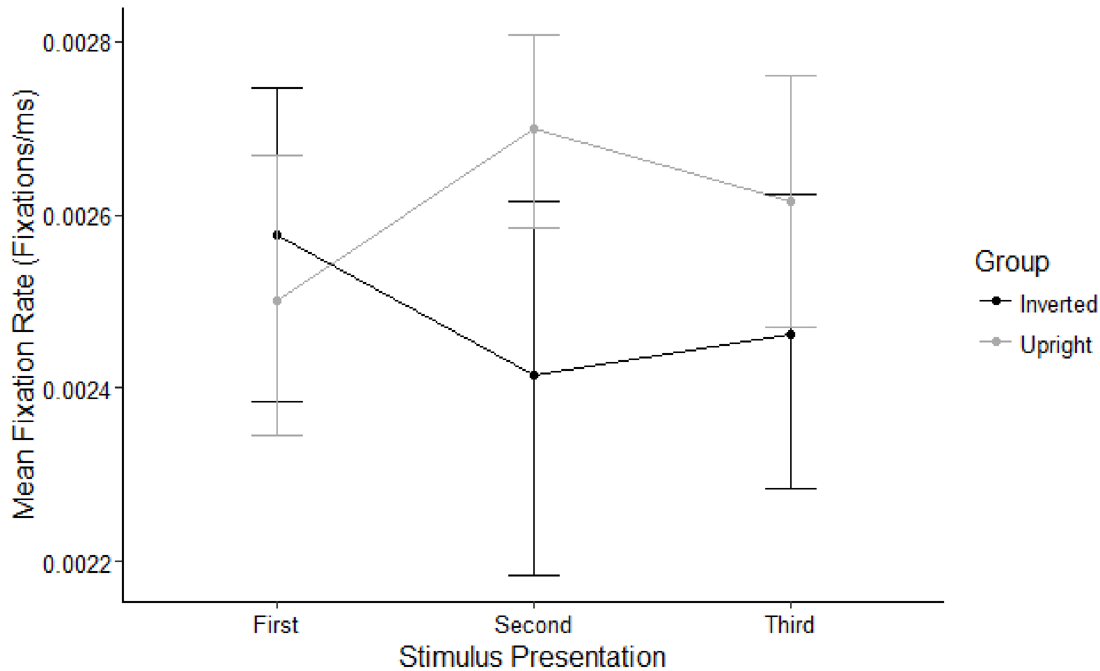

Figure 2. *Phase 1* fixation rate interaction. The graph above displays the interaction between the group and stimulus presentation for fixation rate. Error bars represent a 95% confidence interval.

Table 2. *Phase 1* fixation rate summary.

Effect	df_n	df_d	F	p	η^2
Group	1	50	1.88	.188	0.02
Stimulus Pair	1	50	< 1	.955	0.0004
Group: Stimulus Pair	1	50	3.11	.049	0.03

scan rates during the first stimulus presentation are similar for inverted and upright stimuli. However, the scan rates diverge during the second presentation. Participants in the upright group tend to scan at a higher rate while the participants in the inverted group scan at a slower rate.

Correlations. Since *Phase 1* was also designed to assess natural propensities to scan and consistency in visual scanning behavior, correlations were calculated for the fixation rates within each of the image presentations for the upright and inverted groups and also for consistencies between scan paths. The fixation rates for the inverted group, although not significant, were relatively consistent between image presentation one and two, $r(24) = .31$, $p = .125$, and between presentations two and three, $r(24) = .37$, $p = .062$. However, the consistency drops between presentation one and three, $r(24) = .07$, $p = .744$. Fixation rate consistency for the upright group, also was not significant, but was much more consistent across all of the image presentations: one and two, $r(24) = .28$, $p = .165$, two and three, $r(24) = .34$, $p = .091$, and one and three, $r(24) = .36$, $p = .068$. The fixation rate consistency scatterplots are displayed in Figure 3. Correlations were also conducted for the scan path similarities (Figure 4). Scan paths similarities were much more consistent between comparisons for the inverted group, $r(24) = .40$, $p = .04$, than for the upright group, $r(24) = .11$, $p = .588$.

Phase 2 Analyses

A preliminary analysis was conducted to determine if an order effect was present across the trials. A 2 (Order: Forward vs. Backward) x 10 (Trial) ANOVA was conducted for both the upright and inverted groups. Both analyses yielded no significant order effects, so all subsequent analyses were collapsed by order and only analyzed by the group. Recognition memory trials were reduced to three trial block averages across trials 2 through 10. Trial 1 was omitted from all analyses due to participant confusion about the task requirements. Hence, Trial 1 served, *post hoc*, as a practice trial.

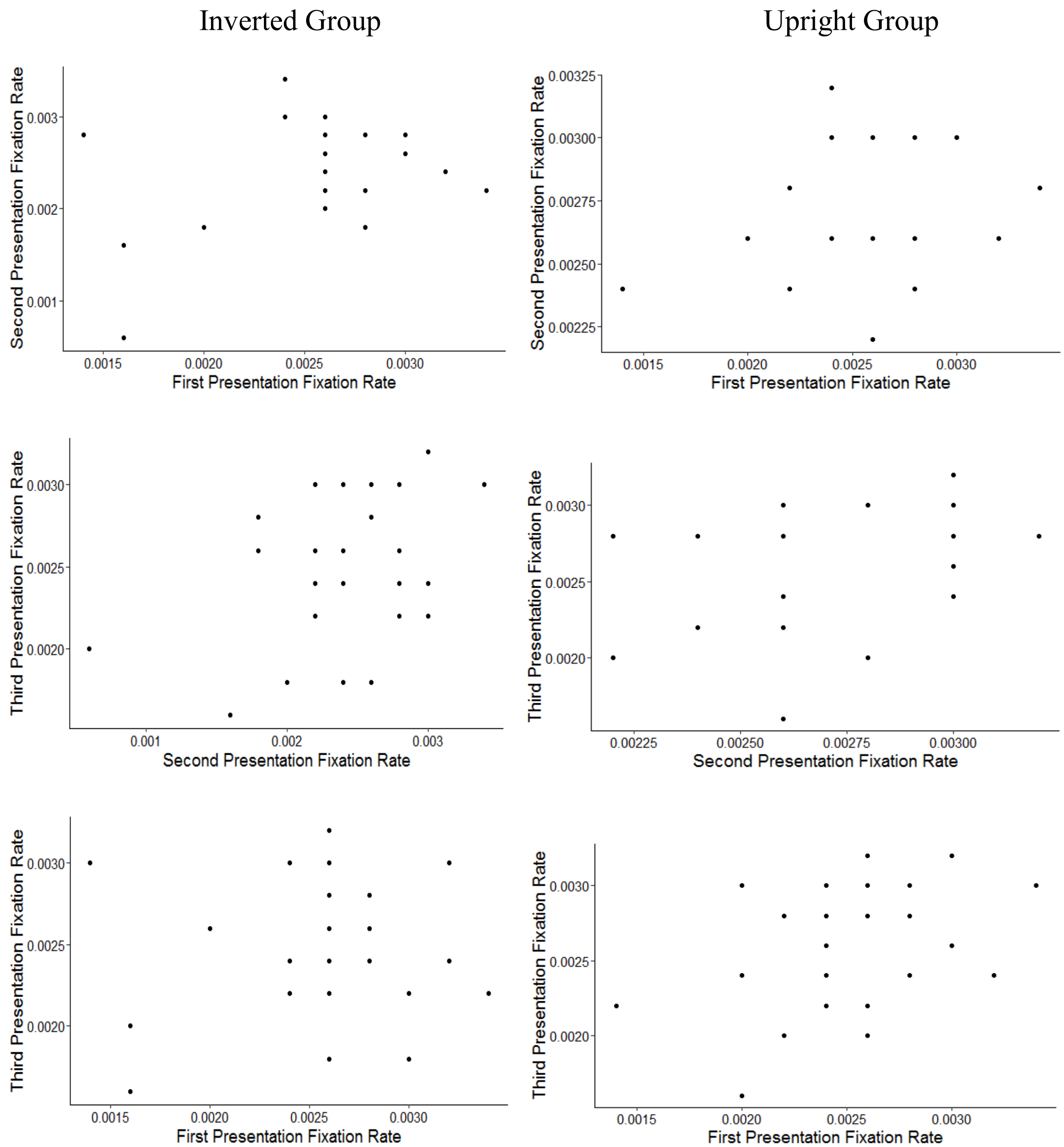

Figure 3. *Phase 1* Fixation Rate Scatter Plots. Scatter plots for the inverted and upright groups to show the consistency in fixation rates across the image presentations.

Response Latency. A 2 (Group) x 3 (Trial Block) ANOVA was conducted for RL. The data are displayed in *Figure 5*. The ANOVA yielded a significant main effect for group, $F(1, 54) = 7.68$, $p = .008$, $\eta^2 = 0.08$, and a main effect for trial block, $F(2, 108)$

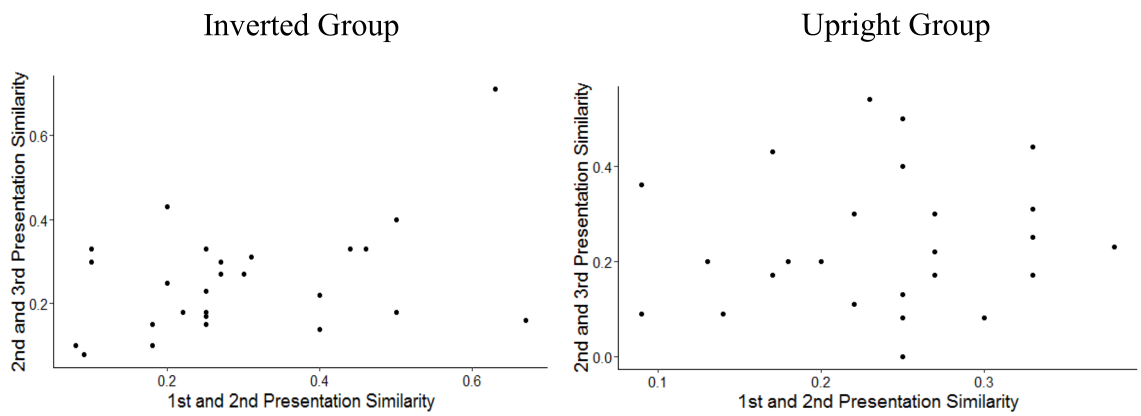

Figure 4. Scan path similarity scatter plots.

= 8.29, $p < .001$, $\eta^2 = 0.05$ (Table 3). Participants in the inverted group ($M = 1636.13$, $SD = 468.03$) had significantly longer RLs than those in the upright group ($M = 1380.21$, $SD = 407.50$) (Figure 5). This finding suggests that there is a difference in how upright and inverted images are processed. It takes longer to recognize and inverted image than an upright one, which is consistent with findings with the FIE. Although the specific mental processes are different from those found in the FIE, it could be argued that participants in the upright group processed the images faster and in a holistic manner, while the ones in the inverted group were required to do a more serial featural search. A *post hoc* analysis, with a Bonferroni correction, was conducted between the trial blocks. A significant difference was found between the first ($M = 1646.67$, $SD = 520.79$) and second ($M = 1489.345$, $SD = 431.54$) trial blocks, $p = .037$, $d_{avg} = 0.33$, and between the first and third ($M = 1402.20$, $SD = 381.17$) trial blocks, $p = .001$, $d_{avg} = 0.54$. There was no significant difference found between trail block two and three, $p = .32$, $d_{avg} = 0.21$. The decrease in RL across trials could be a function of practice and/or the participants developed and become more proficient of the respective strategy for solving the problems.

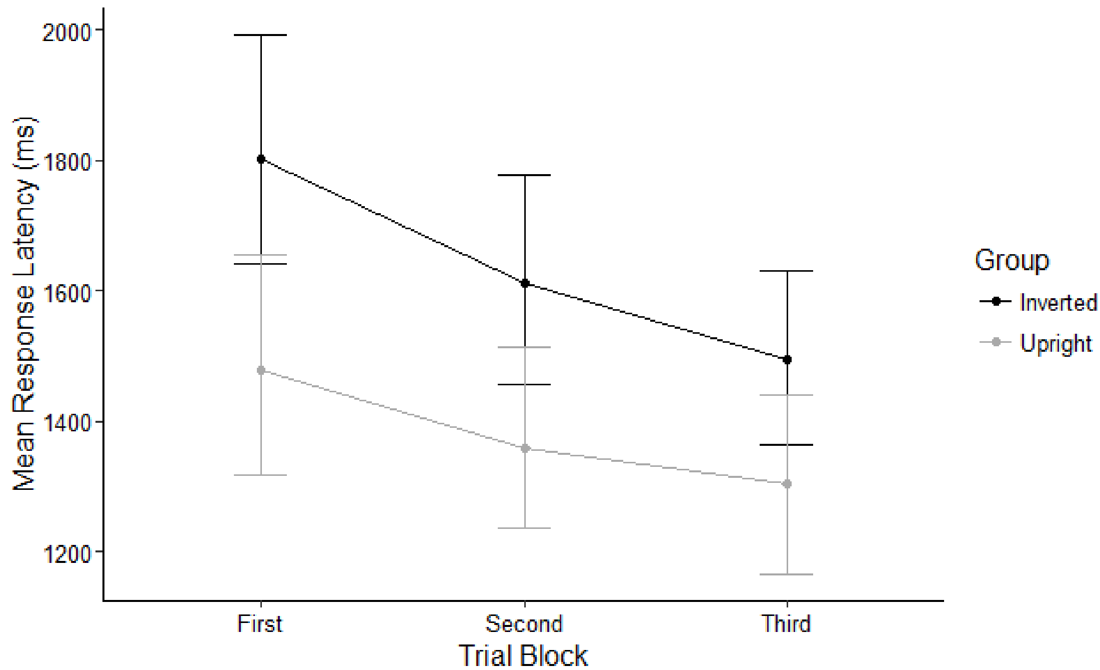

Figure 5. Response latencies across trial blocks. The figure shows the response latencies for both the inverted and upright groups. Error bars represent a 95% confidence interval.

Table 3. *Phase 2* response latency summary.

Effect	df_n	df_d	F	p	η^2
Group	1	54	7.68	.008	0.08
Trail Block	2	108	8.29	< .001	0.05
Group: Trial Block	2	108	< 1	.543	0.004

Fixation rate. Differences in fixation rate were assessed between the groups and the trail blocks via a 2 (Group) X 3 (Trial Block) ANOVA. There was a significant main effect for the trial blocks, $F(2, 84) = 3.83$, $p = .025$, $\eta^2 = 0.04$ (Table 4). A *post hoc* analysis was conducted comparing the trial blocks. While significance was only found between the first ($M = 0.0028$, $SD = 0.00032$) and third ($M = 0.0026$, $SD = 0.00038$)

Table 4. *Phase 2* fixation rate summary.

Effect	df_n	df_d	F	p	η^2
Group	1	42	< 1	.682	0.002
Trail Block	2	84	3.83	.025	0.04
Group: Trial Block	2	84	< 1	.644	0.005

blocks, $p = .024$, $d_{avg} = 0.49$, and not between trial blocks one and two ($M = 0.0027$, $SD = 0.00036$), $p = .844$, $d_{avg} = 0.20$, or two and three, $p = .247$, $d_{avg} = 0.28$, a subtle trend can be seen emerging in the data (Figure 6). The changes in fixation rate across the trials suggest that the participants are making fewer fixations per millisecond as they progress. This finding could be due to strategies developed by the participants to help them succeed in recognizing the target image during the task and this finding is in concordance with the RL data.

Length of first fixation. A 2 (Group) X 3 (Trial Block) ANOVA was conducted on the LFF between the groups and the trial blocks, but yielded no significant results (Table 5).

Phase 3 Analyses

Phase 3 analyses consisted of a series of independent *t*-tests comparing the two groups for fixation rates, number of items recalled, and LFF. There was a significant difference between the upright and inverted groups for fixation rate, $t(45) = 2.56$, $p = .014$, $d = 0.75$. The inverted group ($M = .0028$, $SD = .00046$) fixated the office image at a higher rate than the upright group ($M = .0024$, $SD = .00046$) (Figure 7). This result is

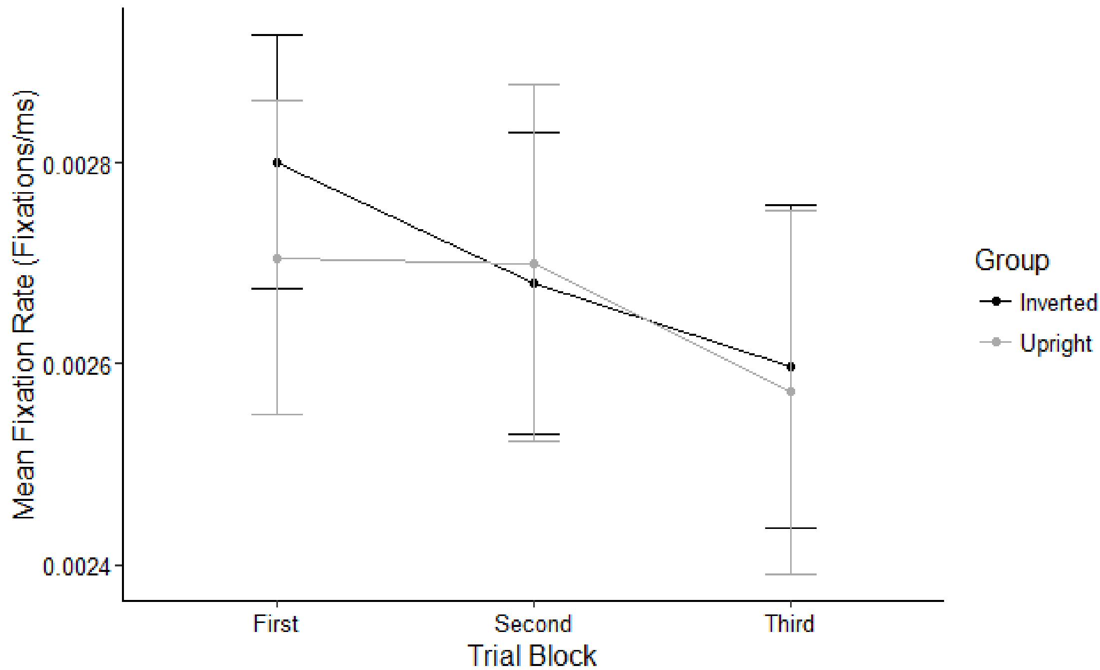

Figure 6. Fixation rate change across trial blocks. Error bars represent a 95% confidence interval.

Table 5. *Phase 2* length of first fixation summary.

Effect	df_n	df_d	*F*	*p*	η^2
Group	1	44	1.91	.174	0.03
Trail Block	2	88	2.21	.116	0.01
Group: Trial Block	2	88	< 1	.784	0.002

counter to initial predictions since it was expected that the inverted group would be processing the images more serially and slowly leading to fewer fixations. The outcome here could be a result of the experimental design and is discussed in greater detail in the discussion section. There were no significant differences between the groups for the number of items recalled, $t(54) = .93$, $p = .36$, $d = 0.26$, or the LFF, $t(45) = -.06$, $p = .57$, $d = 0.18$.

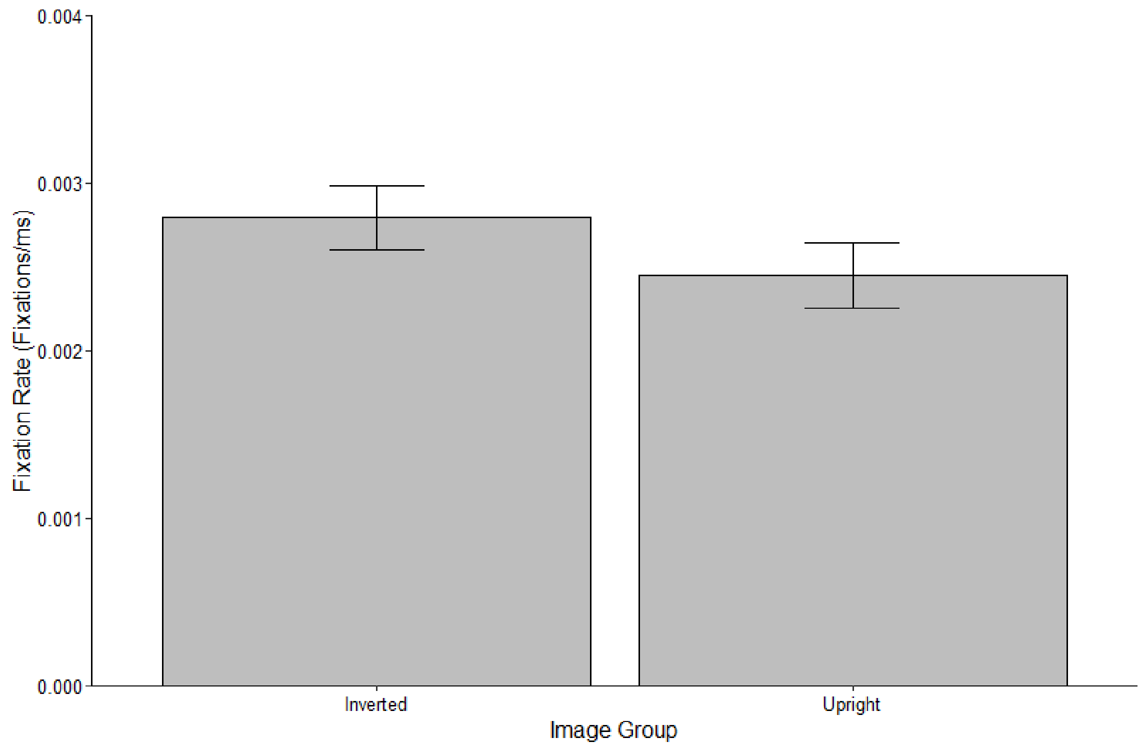

Figure 7. *Phase 3* group fixation rates. Error bars represent a 95% confidence interval.

Correlations for each group were calculated between the number of items recalled and fixation rate. The analyses yielded a significant correlation for both the inverted group, $r(22) = -.45$, $p = .027$, and the upright group, $r(21) = .42$, $p = .048$; however, the direction of the correlations were opposite of each other (Figure 8). For the inverted group, the number of items recall decreased as fixation rate increased. This result suggests that people who were processing the image in a serial fashion had better recall memory, which corroborates our initial predictions. Counter to this finding, participants in the upright group who scanned the image faster, recalled more items from the image. This finding also supports the hypothesis that participants who process the image in a more holistic manner have better recall memory.

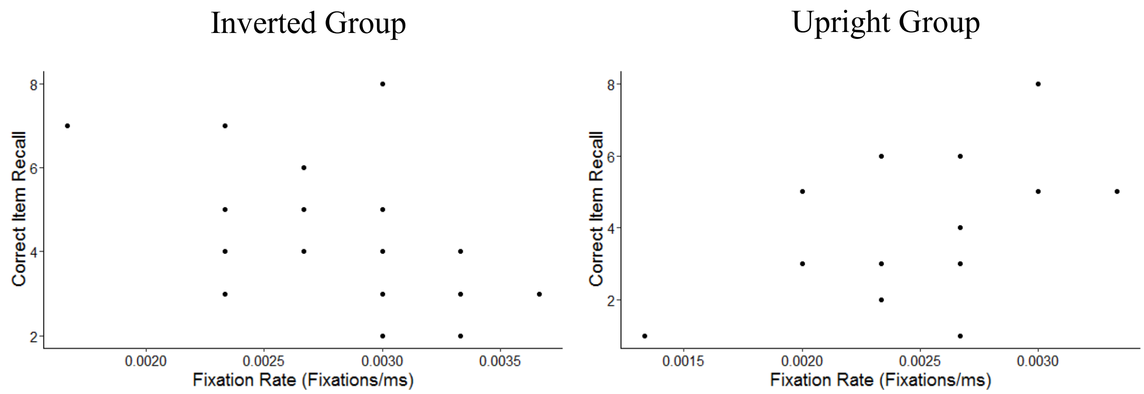

Figure 8. *Phase 3* scatter plots. The plots above show the relationship between fixation rate and the number of items recalled for each of the groups.

DISCUSSION

To summarize, the primary purpose of this study was: (1) to explore different measures of visual scanning to better assess individual differences in perception; (2) to assess differences in visual scanning and RL on inverted and upright images; (3) to assess the relationship between visual scanning measures and RL on subsequent recognition memory test; (4) and to determine if there are differences in the accuracy and number of items recalled after viewing inverted versus upright images. Past research on facial perception and the FIE suggests the processing of faces is hindered when they are inverted, given this, similar effects should be seen when ecological images are also inverted. When face stimuli that are processed holistically, the processing speed (recognition memory) is much faster than when they are processed featurally. When images or faces are inverted the holistic perception process is disrupted and processing tends to become featural. It was hypothesized that this phenomenon would be observed for natural ecological images as has been documented with face stimuli. The results for each phase of the study are now discussed in turn.

Phase 1 Discussion

While the scan path similarities did not differ between the groups, it did seem as though an effect was emerging across the stimulus pairs. Scan paths were more similar for the first stimulus pairing than the second pairing. Aside from increasing the power of the current study, future research should employ more trials across stimuli equated for complexity to better assess scan path trends across the trials. Another manipulation to better assess individual differences, would be to increase the ISI between each stimulus presentation. In the current study, the ISI was only one second. Increasing the ISI could reduce the magnitude (neutralization) of encoded image and therefore the consistency of scan paths could be more evident. Studies such as Harding and Bloj (2010) have a multi-stimulus encoding phase and then a large recognition phase; the variable timing between the image presentations and the large numbers of images might would reduce the magnitude of participant's familiarization of the images. Also, using shorter image presentation times could make individual differences more evident. However, shorter image presentations preclude measurement of visual scanning due to constraints for visual scanning technology; stimulus presentations need to be at least 3 seconds to sample visual scanning. Theoretically, there are at least two models of individual differences that could account for lack of consistency in visual scanning: Sokolov's (1963) comparator model and the Jeffery's (1968) serial habituation model (see Colombo and Mitchell, 2009, for a review). Regardless of model, on repeated presentations an internal representation of the image is formed, and the more salient stimulus components, are encoded first. Hence, on the repeated presentations visual scanning should be reduced. Selective attention would be directed to the less salient stimulus components which presumably have yet to be encoded into memory.

Further research should also be conducted comparing the fixation rates on the *Phase 1* design. The results of this study suggest that there could be a trend in the fixation rates between the upright and inverted groups. The data suggests that, although both groups start at the same rate of fixation, across repeated image presentations, upright image views increase in their rate of fixation while inverted viewers stay about the same or even decrease. Increasing power could allow for these differences to more readily emerge. If they did, it could lend more support to upright images being processed holistically and the inverted images being processed serially. Just as with the scan paths, differences could also become more evident by manipulating the ISI or stimulus presentation time. Also, changes in visual processing could have been better assessed by increasing the number of image presentations. One would expect that on repeated presentations visual scanning should be reduced. Preattentive visual activity would decrease as voluntary selective attention becomes directed to the less salient stimulus components which presumably have yet to be encoded into memory. Although the images were presented only 3 times, more presentations could have resulted in habituation and therefore a better test of changes in visual scanning and processing as a function of repeated presentations.

Individual differences in visual scanning were observed, and it is important to note that visual scanning was a relatively stable measure, particularly for the inverted group. It is the contention of the author that this finding is one of the first stability estimates of visual scanning. If visual scanning is not a stable measure or estimate of individual processing differences one has to question the utility of using visual scanning as a predictor subsequent cognitive outcomes. Granted replication of this result is warranted.

Phase 2 Discussion

The RLs garnered from *Phase 2* were consistent with other similar studies of FIE. Overall, participants did get faster in their RL across the trial blocks. It is the contention of the author that the observed changes in RL is a function of the development and implementation of strategies to solve the recognition memory task; however, participants who viewed the inverted images were slower to respond, on average, than the participants who viewed the upright images. This suggests that inverting the image does interfere with the perceptual process and makes image recognition more difficult. However, since there were no differences between the groups in visual scanning on this part of the study, it is difficult to explain this interference. The other phases do suggest that there is a difference in how the images are scanned, so it could be an issue, not only with the study being under powered, but related to the type of stimuli used.

While LFF did not show any significant differences, there was a difference in fixation rate for both groups across the trials. Participants made fewer fixations per millisecond across the trial blocks. This finding could be part of the strategy they used to help with image recognition and could be the reason differences were not seen between the groups. If the participants in the upright group developed a strategy of encoding specific contexts or holistic aspects of the image, it seems that fixation would decrease. For example, several participants said that they would just pay attention to the type of car that might be present in the scene to aid their recognition in the task. More research should be conducted using different stimuli (such as more natural scenes) to see if similar results are found.

Phase 3 Discussion

Phase 3 analyses yielded some intriguing and unexpected results. First, it was hypothesized that participants in the inverted group would be able to recall more items from the office image than the ones in the upright group. However, the analysis did not yield any significant difference between the two groups. This pattern could be because they actually are recalling about the same number on average, but the size of the effects suggests that it could be an issue with power and the sample size or due to the novelty of the office image. There is a significant difference in the fixation rate between the two groups, with the inverted participants scanning faster than the upright group. This type of trial would be another area in which further research is needed to tease out if difference is due to the type of stimulus or the novelty of the stimulus.

Conclusion

One of the primary goals of this research was to provide further evidence of individual differences in the perceptual processing of ecologically valid stimuli, and by investigating comparatively and closely the differences in derived measures of visual scanning. The visual scanning findings do provide some guidance and clarification of how visual scanning measures can better represent perceptual processes that are necessary for recognition memory. Furthermore, there is a difference in the way people perceive and recognize upright and inverted images, similar to those found in research on the FIE.

This study also demonstrated changes in visual scanning across trials; a decrease for both the inverted and upright groups. This finding could be interpreted as support for the two-process stage models of perception proposed by Neisser (1976) and Gibson (1966) decades ago. Borrowing from Mitchell's (2005) model of visual discrimination learning, it could be argued that the rate of visual scanning changed as a function of changes in the allocation of attention during the image presentation. Initially superordinate allocation of attention was guided by preattentive processes which encompassed stimulus feature orientation and detection; and the subordinate allocation of attention was guided by the perceiver selectively attending to stimulus features and constructing a perceptual object(s). What is lacking, is an assessment of the kinds of cognitive structure(s) needed to form said objects. It is the contention of the author that cognitive processing begins with perceptual behavior (Cooper and Regan, 1982) and if cognitive science is to advance, the measurement and development of individual perception differences are necessary to advance cognitive theory. Launching out new directions in the study of visual scanning appears to be a promising venue.

REFERENCES

Borg, A., & Sariyar, M. (2016). RecordLinkage: Record Linkage in R. R package version 0.4-9. Retrieved from https://CRAN.R-project.org/package=RecordLinkage.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*(1), 28-38.

Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development, 65*(5), 1243-1261.

Cashon, C. H., & Cohen, L. B. (2003). The construction, deconstruction, and reconstruction of infant face perception. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood* (pp. 55-68). New York: NOVA Science.

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory, 92*, 225-234.

Cooper, L. A., & Regan (1982). Attention, perception, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. New York: Cambridge University Press.

DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience, 5*, 219-226.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 1-17.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin, Boston.

Harding, G., & Bloj, M. (2010). Real and predicted influence of image manipulations on eye movements during scene recognition. *Journal of Vision*, *10*(2), 1-17.

James, W. (1890). *The principles of psychology*, *Vol I*. Henry, Holt, & Company, New York.

Jeffrey, W. E. (1968). The orienting reflex and attention in cognitive development. *Psychological Review*, *75*, 323–334.

Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience and Biobehavioral Reviews*, *50*, 169-179.

Lawrence, M. A. (2015). ez: Easy analysis and visualization of factorial experiments. R package version 4.3. Retrieved from https://CRAN.R-project.org/package=ez.

Lui, Y. M., Bolme, D., Draper, B. A., Beveridge, J. R., Givens, G., & Phillips, P. J. (2009). A meta-analysis of face recognition covariates. *IEEE Third International Conference on Biometrics: Theory, Applications and Systems*, 1-14.

Mathôt, S., & Theeuwes, J. (2011). Visual attention and stability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 516–527.

Mitchell, D. W. (2005). A heart rate model of visual discrimination learning. Paper presented at the annual Association for Behavior Analysis Conference, Chicago, IL.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLEARN: A two-process theory of infant face recognition. *Psychological Review, 98*(2), 164-181.

Nakabayaski, K., & Liu, C. H. (2014). Development of holistic vs. feartural processing in face recognition. *Frontiers in Human Neuroscience*, *8*, 1-5.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. W. H. Freeman, New York.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, & P. U. Tse (Eds.), *Progress in Brain Research* (Vol. 155, pp. 23-36). Amsterdam: Elsevier B. V.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *TRENDS in Cognitive Sciences*, *11*(12), 520-527.

Phillips, M. H., & Edelman, J. A. (2008). The dependence of visual scanning performance on search direction and difficulty. *Vision Research*, *48*, 2184-2192.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10-15.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/.

Rakover, S. S. (2013). Explaining the face-inversion effect: the face-scheme incompatibility (FSI) model. *Psychonomic Bulletin & Review*, *20*(4), 665-692.

Reynolds, J. K. & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instructions. *Applied Cognitive Psychology, 6,* 279-292.

Richler, J.J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin, 140(5), 1281-1302.*

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012) Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology, 3*, 1-6.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1-3), 157-173.

Shapiro, P. N., & Penrod, P. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*(2), 139-156.

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual review of physiology*, *25*, 545-580.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A, 20*(7), 1407-1418.

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

Treisman, A., Vieira, A., & Hayes, A. (1992). Automatic and preattentive processing. *American Journal of Psychology, 105,* 341-362.

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefèvre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision, 10*(5), 1-13.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software, 21*(12), 1-20.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141-145.

Yin, R. K. (1970). Face inversion by brain-injured patients: A dissociable ability? *Neuropsychologia*, *8*, 395-402.

APPENDIX

Missouri State University Consent of Participation – Benjamin Graves - 2016
Infant Perception and Learning Laboratory

This study is part of the Missouri State University Psychology Graduate Program designed to give us more information and to fulfill a thesis requirement for Benjamin Graves. The following information is provided so that you can decide whether you wish to participate in this study. If you agree to participate, we will observe your visual responses to a series of slides of natural environmental scenes. One of the members of the research lab should have explained the purposes and procedures of the study to you, and will answer any questions you might have. Please be assured that if you agree to participate, you are free to withdraw from the study even after you have signed this consent form. If you wish to withdraw, simply stop any on-going task and tell the research staff you wish not to continue. Should you decide to terminate the research session; all data pertaining to you that have been collected will be destroyed.

Since it is our policy to protect the confidentiality of all our participants, your name will not be included in any data analyses, subsequent publication or presentations related to this research study. All raw data collected during this study will be identified only by code-number to insure confidentiality of the information collected.

If questions arise after you have left the research laboratory, feel free to give D. Wayne Mitchell, Ph.D. a call at 417-836-6941 or at waynemitchell@missouristate.edu. We do not anticipate any risk to you as a result of participating in this study, but it is unlikely that this study will provide you with any direct benefits. Your participation will, however, make an important contribution to our scientific knowledge, and we very much appreciate your cooperation.

In addition, we would appreciate your filling out the attached demographic sheet so we can document the characteristics of our participants. Any of the questions you feel uncomfortable about answering, please feel free to leave blank. As with the raw data collected, this information will be entered into our computer system and only identified by code-number to insure confidentiality.

I have read the above description of the study and I agree to participate.

Participant's Name (please print):____.

Participant’s Signature:____.

Witness’s Signature: ____.

Date: ___________________________________.

DEMOGRAPHIC INFORMATION SHEET – Benjamin Graves 2016

Participant's Name: ____.

1. Date of Birth ___________________________.

2. Gender _______________________________.

3. Major ________________________________.