

The Pennsylvania State University
The Graduate School
College of Health and Human Development

**EFFECTS OF PLACEMENT OF HUMAN FIGURES WITHIN NATURALISTIC
SCENES (PHOTOGRAPHS) ON INITIAL VISUAL ATTENTION OF YOUNG ADULTS**

A Thesis in
Communication Sciences and Disorders

by
Alyssa Kneisly

© 2011 Alyssa Kneisly

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2011

The thesis of Alyssa Kneisly was reviewed and approved* by the following:

Krista Wilkinson
Professor of Communication Sciences and Disorders
Thesis Adviser

Janice C. Light
Distinguished Professor of Communication Sciences and Disorders

Carol A. Miller
Associate Professor of Communication Sciences and Disorders

Kathryn D.R. Drager
Associate Professor, Professor-in-Charge of the CSD Graduate Program

*Signatures are on file in the Graduate School.

Abstract

Purpose: There are many systems and devices available for use in augmentative and alternative communication, there are many systems and devices available for use in intervention. This research focused on the Visual Scene Display (VSD) in which language concepts are embedded within broader scenes depicted within photographs or other schematic images. Currently, naturalistic human figures are not frequently used in VSDs. This research examined how human figures attract visual attention during a short viewing period.

Method: Twenty undergraduate and graduate students viewed one of three image sets (each set consisting of three images) over a short viewing period of five seconds per image. The participants were given minimal verbal instruction to provide a "free viewing" setting. Infrared eye-tracking technology was used to record the participants' viewing patterns. For each image, latency to first view and duration of view for each element were measured.

Results: The human figures were strong attractors of visual attention even when they competed with other elements, larger elements, pseudo-human figures and implied motion for visual attention. Placement of the human figure had little effect on these initial viewing patterns.

Conclusions: Human figures are strong attractors of visual attention regardless of size or placement within an image. These findings have implications for the programming of VSDs in terms of inclusion of human figures and the use of contextually rich images.

Table of Contents

List of Tables	vi
List of Figures	vii
Acknowledgements	viii
Introduction	1
Human Figures and Communication Sciences and Disorders	1
Visual Scene Displays as an Alternative to Traditional Formats.	2
The Potential Role of Human Figures in VSDs	4
Literature Review	7
Historical Background and Seminal Works	7
Human Figures as a Focus of Visual Attention	10
Summary and Purpose	12
Method	14
Participants	14
Materials	16
Criteria for Selecting Images	16
Experimental Conditions	17
Coding of Selected Images.	19
Kids-dog	20
Family-statue	21
Man-bottle	22
Data Collection	24
Dependent Measures	25
Results	28
Kids-dog	28
Family-statue	30
Man-bottle	32
Discussion	34
Kids-dog	34
Family-statue	36
Man-bottle	39
Summary and Conclusion	41
Applying Findings to the Field of Communication Sciences and Disorders	42
Limitations and Future Research	42

References.....	46
Appendix A: Figures A1-A3.....	48
Appendix B: Figures B1-B3	51
Appendix C: Figures C1-C3	54
Appendix D: Figures D1-D3.....	57
Appendix E: Figures E1-E3	60
Appendix F: Figures F1-F3.....	63
Appendix G: Figures G1-G3.....	66
Appendix H: Figures H1-H3.....	69
Appendix I: Figures I1-I3	72

List of Tables

Table 1.Participants Used in Set One	15
Table 2.Participants Used in Set Two.....	15
Table 3.Participants Used in Set Three	16
Table 4.Percent of Image Occupied for Kids-dog	21
Table 5.Percent of Image Occupied for Family-statue	22
Table 6.Percent of Image Occupied for Man-bottle	23
Table 7.Number of Participants' Data Used Per Image	25
Table 8.Split-half Analyses of Selected Image Conditions	26

List of Figures

Figure 1. Stimuli images listed in sets	18
Figure A1. Kids-dog with human figures absent	47
Figure A2. Kids-dog with human figures on left	48
Figure A3. Kids-dog with human figures on right.....	49
Figure B1. Family-statue with human figures absent	50
Figure B2. Family-statue with human figures on left	51
Figure B3. Family-statue with human figures on right.....	52
Figure C1. Man-bottle with human figure absent	53
Figure C2. Man-bottle with human figure on left.....	54
Figure C3. Man-bottle with human figure on right.....	55
Figure D1. Kids-dog: Element coding with human figures absent.....	56
Figure D2. Kids-dog: Element coding with human figures on left.....	57
Figure D3. Kids-dog: Element coding with human figures on right	58
Figure E1. Family-statue: Element coding with human figures absent	59
Figure E2. Family-statue: Element coding with human figures on left.....	60
Figure E3. Family-statue: Element coding with human figures on right.....	61
Figure F1. Man-bottle: Element coding with human figure absent	62
Figure F2. Man-bottle: Element coding with human figure on left.....	63
Figure F3. Man-bottle: Element coding with human figure on right.....	64
Figure G1. Kids-dog: Latency to first view	65
Figure G2. Kids-dog: Time spent viewing elements relative to total viewing time	66
Figure G3. Kids-dog: Ratio of visual attention to elements	67
Figure H1. Family-statue: Latency to first view	68
Figure H2. Family-statue: Time spent viewing elements relative to total viewing time	69
Figure H3. Family-statue: Ratio of visual attention to elements across orientation.....	70
Figure I1. Man-bottle: Latency to first view	71
Figure I2. Man-bottle: Time spent viewing elements relative to total viewing time.....	72
Figure I3. Man-bottle: Ratio of visual attention to elements	73

Acknowledgements

I would like to thank my advisor, Dr. Wilkinson for her enthusiasm and giving me the impetus to choose the thesis format for my research. My other committee members, Dr. Light and Dr. Miller, were invaluable in their insight in the development of this thesis and helped to form an incredibly well-rounded committee. I also extend many thanks to my co-researcher, Chihui Yong, who knows just about all there is to know about the ISCAN software and hardware.

Introduction

The familiar aphorism "A picture is worth a thousand words," has never held more true than when applied to the field of visual attention research. Rather, millions, and perhaps billions of words have been employed in the examination and description of human visual attention patterns and pictorial stimuli. The purpose of this research is to add to the current knowledge base by examining underlying factors affecting eye gaze patterns, or "scanpaths" (Noton & Stark, 1970, p.933) recorded by eye-tracking technology, in order to ask: "How are human eye gaze patterns affected when the placement and inclusion of human figures within a scene are manipulated?" This research seeks to replicate three original conditions of images analyzed by Wilkinson and Light (2011) in addition to two new manipulations involving orientation and the absence of human figures to study the way in which people viewing photographs over a short period of time (five seconds) attend to the human figures within them. The long-term goal of the research is to use this knowledge to help optimize the construction of visual aids that are used to foster communication in children with disabilities. The research focuses particularly on human figures for a number of reasons to be detailed in this introduction.

Human Figures and Communication Sciences and Disorders

Augmentative and alternative communication (AAC) represents one of the many areas of study that fall under the broad umbrella of communication sciences and disorders (CSD) and into the realm of speech-language pathologists (SLPs) and related professionals. AAC can be described as "a set of procedures and processes by which an individual's communication skills (i.e., production as well as comprehension) can be maximized for functional and effective communication" (American Speech-Language-Hearing Association, 2002, p. 2). The systems used to maximize communication can take the form of "light tech" or "high tech" systems.

"Light tech" may be as simple as a piece of paper containing the letters of the alphabet. "High tech" systems are generally computer-driven and may be dedicated, or non-dedicated. An example of a dedicated system is the DynaWrite (produced by DynaVox Mayer-Johnson) as it can only be used for communication. The device is not specifically designed to be used for other activities that may be performed on a computer such as word processing or accessing the internet. Conversely, non-dedicated systems, such as the DynaVox VMax, also produced by DynaVox Mayer-Johnson, provide users with a high-tech system that works in conjunction with a traditional computer setup, in this case, the Windows computing platform. This system differs from a dedicated device because it allows the user to access the internet and operate programs such as Microsoft Office in addition to functioning as a communication device.

Currently, a wide variety of these and other AAC systems exist that use many different pictures, symbols, and images to represent language concepts. Historically, these concepts have been presented to users in a grid format, with symbols lined up into rows and columns for presentation and selection.

Visual Scene Displays as an Alternative to Traditional Formats. An alternative to the row-column presentation of language concepts is the Visual Scene Display (VSD). These are a fairly new but potentially useful addition to the AAC framework. Blackstone (2004) describes VSDs as scenes that "are used as an interface to language and communication" (p. 3). More specifically, they "portray events, people, actions, objects and activities against the backgrounds within which they occur or exist" (Blackstone, 2004, p.3). In addition to presenting a fixed image, VSDs may also contain "intelligent agents," or elements that have been animated so that they may interact with and be moved throughout the scene (Blackstone, 2004).

Drager and her colleagues (2003) argued that the information provided by the scene provides contextual support for both communicators and their communicative partners. Thoughts and ideas may be initiated and transferred more easily when communication is augmented by the rich, nonlinguistic context images can provide. In turn, the contextual support provided may also decrease the communicative demand on both the communicator and the communicative partner thus facilitating conversation.

Visual scene displays can range from symbolic and generalized "cartoon"-like images such as Mayer-Johnson DynaVox Boardmaker line drawings to highly personalized digital photos representing an individual's unique experiences. Ideally, users of AAC could take and send pictures to use to communicate just as cell phone users are able to take and send images to their friends and family. By accessing currently available technology, AAC users' systems could be made more relevant by giving them the power to communicate about recent events without having to wait for programming to make them accessible.

Why is it potentially important to consider dimensions of displays such as the personalization, context, and so forth? One reason is that it takes into consideration the linguistic needs of children such as contextual support. In a descriptive study, Light, Page, Curran and Pitkin (2007) asked children to design communication devices. The children identified roles for the devices to fulfill such as companionship, humor, play and artistic expression (Light, Page Curran & Pitkin, 2007). In addition, they identified bright colors, interactive and moving parts, and personalization as important aspects that were desirable in their communication devices. Considering how well communication aids encompass these dimensions, Light, Drager and Nemser (2004) found that AAC systems in their systematic analysis lacked many of the characteristics of the popular children's toys that they sampled. Overall, the AAC systems' dull

colors, lack of interactive and moving parts, and child-friendly themes differentiated them from the popular toys (Light, Drager & Nemser, 2004). Their study indicates that there is currently insufficient "consumer wants and needs" research behind the commercial development of AAC devices. By further researching the qualities of successful toys, companies could create devices that are desirable to the individuals that require AAC to communicate.

Furthermore, considering the diverse challenges faced by individuals who require AAC, it is the priority of SLPs and others to create the most efficient systems possible. Users of AAC are as diverse as the types of AAC they require. One of the most difficult parts about developing AAC systems is designing them with the potential to be generalized over as broad a set of needs as possible. Often, individuals requiring AAC must have systems that accommodate for visual processing, fine and gross motor function and cognitive load. In addition, these AAC systems must be adaptable to any number of real-life situations from the classroom to a day at the beach. Clearly, the challenge in providing appropriate AAC systems is great.

The Potential Role of Human Figures in VSDs. I opted to examine the role of human figures as areas of interest within VSDs in order to be better able to create visual scene displays that are user-friendly and socially engaging (as opposed to socially isolating) and are able to clearly communicate messages and minimize confusion. One reason humans have been identified as a key feature of the current study is because researchers have already focused on visual characteristics such as composition luminance, color and size, (Niekamp, 1981; Vincent, Baddeley, Correani, Troscianko & Leonards, 2009; Chen & Zelinsky, 2006) and the physiological characteristics and limitations of individuals (e.g. motor coordination and motor processing) (Castelhano & Henderson, 2008; Tatler & Vincent, 2009). Determining how specific elements featured in images attract visual attention could allow SLPs to maximize AAC

systems in order to make the image-based presentation of language concepts as attractive and easy to understand as possible. This would benefit not only the individuals who require AAC, but their communicative partners as well. These findings could then be considered along with the findings from Light, Drager and Nemser (2004) as well as Light, Page, Curran, and Pitkin (2007) as a foundation of research to begin developing more consumer-and-user-friendly AAC systems.

Another benefit of the use of human figures in VSDs and AAC is that they can be socially engaging to users and potential communication partners. Due to cultural and ethnic diversity between users of AAC and their communicative partners, some abstract symbols (and even the English language) may not always be clearly understood. The inability of potential communication partners to understand certain symbol systems may result in the AAC user becoming socially isolated in certain situations. By making humans a central feature in visual communication, the act of communication between an AAC user and a communication partner can become a shared experience in which both can engage in the visual scene and relate in some way to the image presented. Instead of using a system by which the AAC user can interact with others, the use of VSDs now gives others a way with which they too can interact. For example, a child using a traditional symbol-based AAC system may not be able to communicate well with his peers if his peers are unable to understand and operate his AAC device. The symbols may be confusing and understanding how to interact with the device may not be intuitive to a young child (as opposed to the adult that created it). In this situation, a VSD could benefit the child by providing a context-rich and relatable scene, say an image of a child playing on a beach on vacation. This natural and personal image would offer more context to a communication partner than a set or grid of symbols depicting sand, a beach or water. In addition, the communication

partner would be better able to talk about the child's vacation directly as opposed to having to ask questions to clarify and understand the underlying "vacation" message. By using natural, relatable images of humans, the AAC systems of individuals who require AAC will also become more accessible to their communicative partners and may help to reduce confusion.

In addition to informing us about the role of human figures in scenes, the research also aims to further understanding of the constructs behind human eye gaze patterns.

Literature Review

This review of the literature begins with the two works that are considered seminal works in the field of eye gaze research. In addition, the role of human figures as salient features will be examined. For the purposes of this research study, salience is defined as the visual importance denoted by the amount (in number and length of views) given to an element (or object) within a visual scene.

Historical Background and Seminal Works. In 1935, Guy Thomas Buswell published a qualitative study that is considered to be one of the two seminal works in the study of gaze movement in pictorial stimuli. Through his work, Buswell attempted to uncover the psychology of what affects an individual's visual attention to particular aspects of an image. "How People Look at Pictures" is an incredibly thorough work in which Buswell conducted a variety of experiments that are characterized by a complexity which certainly distinguish it from more modern studies. The stimuli were varied, including photographs, paintings, and abstract art. Perhaps most interesting about this study is how similar the technology used to record eye gaze patterns in 1935 is to the technology used today. Buswell utilized a six-volt filament lamp, a series of mirrors, a camera lens, two prisms, and a moving kinetoscope film to record eye movements by the movement of a light reflected off of the participants' corneas. The modern technology used in this study achieved the same effect in a dramatically simplified form, using an infrared camera, a computer, and an eye-gaze software program.

The results of Buswell's (1935) work consisted largely of individual participants' gaze patterns superimposed over the corresponding visual stimuli and density plots of all participants' gaze patterns superimposed over the corresponding visual stimuli. The strength of presenting data in these forms is that that it allows for contrast between the figures both individually and

grouped by characteristics such as age, background, and cultural bias. In total, Buswell reported using about 200 participants, the majority of which were college-age students.

Another less orthodox (which appears to have been abandoned in modern research) form of presenting eye fixation data was achieved by superimposing a numbered grid over the picture stimuli and noting the percentages of fixations that corresponded to each quadrant. While Buswell excelled in presenting the eye gaze data visually, his data tables represent what is arguably the weakest component of his study. It is particularly important to note that virtually no statistical analyses were performed in a study that presents an incredible amount of raw data. The data tables that Buswell did present merely stated fixations in terms of amount (number of fixations), percentage of fixations, or length of fixations in seconds. The only actual analysis Buswell offered was vaguely qualitative and consisted of describing his data in written form and venturing conservative explanations. In his conclusion, Buswell states that "perceptual processes of the person" are the primary motivating factors in determining eye fixations (Buswell, 1935, p. 144). He also states that "students of art" are the primary audience for his study (Buswell, 1935, p.144). It might be implied that the lack of statistical analyses is attributable to Buswell's target audience, but it is unknown if the absence of statistical analyses was deliberate.

The second seminal work in the field was published by Alfred L. Yarbus in 1967. In contrast to Buswell (1935), Yarbus' work, "Eye Movements and Vision", was directed toward "students and researchers in the fields of biophysics, physiology, medicine, psychology" (Yarbus, 1967, p.ix). Furthermore, Yarbus stated that he believed his research could potentially be of interest to commercial advertisers as well. Yarbus' scientific perspective is made clear in his attention to not only physical behaviors of the eyes such as saccadic movements, but in his discussion of the physiological structure of the eyes as well. It is this commonality that makes

Yarbus' work more connected overall as the role of rapid involuntary eye movements known as saccades is discussed alongside each experiment. By combining physiology with his findings, Yarbus not only justifies his findings but also seeks plausible physiological explanations for them. Like Buswell, Yarbus also sought to understand the visual relationship between a viewer and an image as well as to identify factors that can manipulate the visual relationship.

While Buswell did not examine voluntary and involuntary eye movements, Yarbus capitalized on their presence throughout his study. Yarbus' ability to separate physical movements and mental processing (such as being given directions while looking at a stimulus) made his work one of the most famous examples of "top-down" processing. One reason for this may be the difference in eye tracking technology between the two studies. Yarbus (1967) took a much different approach from Buswell by eliminating the need for the large and rather Rube Goldberg-esque setup by reducing the equipment to a small mirror attached to a suction cap that was fitted over participants' eyes. By moving the recording equipment to the participants' eyes, he may have been able to capture more accurately the saccadic eye movements that are featured so prominently in his work. Unfortunately, the nature of the cap forced participants' eyes to remain open which limited experiments to about five minutes in length to prevent the participants' eyes from becoming painfully dry

One item in particular that is missing from Yarbus' work is a description of his participants. In fact, the participants themselves are rarely mentioned except when referred to by their patterns on a given task. Their confidentiality is so highly maintained that Yarbus did not appear to report any information on them in his study. Yarbus occasionally uses letters of the alphabet to present a particular set of data, but this labeling is inconsistent and much of the data are anonymous. Another weakness of the study is that Yarbus did not organize data by

participant characteristics as Buswell did. This omission raises a question of whether or not Yarbus himself maintained a high level of confidentiality or did not himself know which data set belonged to which participant. In addition, the small numbers of participants that appeared throughout the studies (in some cases only one-to-two participants), make it virtually impossible to know the total number of individuals that participated.

Despite Yarbus' scientific focus, his research also lacked statistical analyses. Though he provides thorough discussion on the physiology of the eye, data from the eye movements are either presented in pictorial form by scanpath patterns or in tables that list the length of saccadic eye movements. However, this exclusion is even more unusual in Yarbus' work because it clearly exhibits a much stronger scientific focus and is directed toward a scientific audience.

Human Figures as a Focus of Visual Attention. While there is not currently a large body of research specifically devoted to the study of human figures in visual scene displays, two recent studies have examined the effect of visual attention on human figures. Humphrey and Underwood (2010) examined human figures within visual scenes and their effect on memory. They tested 100 pictures with 15 participants and found that there were an increased number of saccades to scenes that contained human figures compared to scenes that did not. The researchers also found the number of human figures, the size of the human figure's region of interest (ROI), and the distance of the human figure from the participants' initial fixation all failed to have as strong an influence on visual attention as the presence of the human figure in the scene. In addition, the scanpaths of participants who viewed images containing human figures were more similar than the scanpaths to images viewed that did not contain human figures. The conclusions drawn by Humphrey and Underwood (2010) are especially relevant to

this study because they provide evidence for the significance of human figures within visual scenes.

Another group of researchers, Crouzet, Kirchner and Thorpe (2010), also studied human figures focusing more on saccades made toward faces. Their experiment first involved showing participants one of three types of images: faces, vehicles and animals. Participants were then given a “target category” for which they were instructed to look at first when presented with a pair of images. In addition, a set of “neutral distracter” images was used which featured subjects such as rocks, ice and road signs (Crouzet, Kirchner, & Thorpe, 2010). Each pair of images was presented at 400ms in order to minimize extended visual exploration of the image while still providing an opportunity to record several saccades and a brief scanpath. An interesting result of this study was the finding that across the three categories, faces were visually attended the fastest and with the greatest accuracy. The speed at which participants made saccades to the faces were the fastest, at 110ms followed by animals at 120ms, and vehicles at 140ms.

The second part of the experiment involved only the face and vehicle categories of images in an anti-saccade procedure. Participants were again presented with two side-by-side images and asked to visually attend the images belonging to a pre-specified category, for example vehicles, while visually neglecting (making no saccades) to the other category. The results of this experiment showed that participants were unable to keep from viewing faces (even when instructed not to) and in fact made earlier saccades to the faces than the vehicles they were supposed to be targeting. This inability to control early saccades was also noted by Nummenmaa, Hyönä, and Calvo (2009) in their study of the emotional content of stimuli. In addition, the researchers concluded that participants were both faster and more accurate when the target image (regardless of category) appeared on the left as opposed to the right. This finding

was also supported in their third experiment, which presented pairs of images in differing left, right, and vertical orientations. Interestingly, changes in vertical orientation of the images were not found to have the same effect seen as changes in left/right orientation.

Fletcher-Watson, Findlay, Leekman and Benson (2008) performed a similar study by presenting participants with a pair of images. Each pair of images contained a natural scene featuring a human figure as well as a natural scene with no human figure. They found that, in a free viewing condition, 67% of participants visually attended the scene containing the human figure while the scene with no human figure was visually attended by only 26% of participants. This difference in visual attention was also supported by the participants' first fixations which were "significantly more" to scenes containing human figures (p. 576). The strong visual salience of human figures led them to conclude that human figures (faces and bodies, specifically) "are subject to special perceptual processing" (p. 571).

Summary and Purpose

This research aimed to understand how visual attention within a short viewing period might be affected by a systematic replication of manipulations (altered orientation and absence of human figures) of three images first analyzed by Wilkinson and Light (2011). I asked, 1) "How rapidly and for how long do human figures capture attention within naturalistic photographs?" and 2) "In what ways does the orientation of human figures within an image (left vs. right) affect the duration of visual fixations to human figures and other elements?" While the current literature (Crouzet et al., 2010; Fletcher-Watson et al., 2008; Humphrey & Underwood, 2010) suggests that the presence of humans in visual scenes strongly attract visual attention, my study aims to better our understanding of humans as salient visual elements.

The two manipulations for my study were chosen based on the prominent theories of visual attention: 1) manipulating the pictures so that for each image there are two conditions where the human figures appear oriented to the left and also to the right, 2) manipulating the pictures so that for each image there was a condition in which the human figures were digitally removed. The reasoning behind these manipulations was to probe for possible differences in orientation such as those previously explored by Nummenmaa, Hyönä, and Calvo (2009). I hypothesized that the following outcomes would occur: 1) Human figures would attract attention rapidly (as measured by latency to first view) when compared to the non-human elements in visual scene displays; 2) Visual attention to objects within the image would differ based on the orientation of human figures within the image (left vs. right); and 3) The fixations analyzed would reveal a viewing pattern indicating a bias related to the presence of the human figures.

Method

Participants

Participants were recruited from the student body at the Pennsylvania State University. The Human Participants Institutional Review Board of the Pennsylvania State University had already approved this research as an extension of Dr. Wilkinson's pre-existing projects. A total of 24 students participated; there were 14 females and 10 males. Due to calibration errors and inconsistencies, the data from seven participants were excluded from this study. Of the 24 total participants, four participants' data were removed from the study completely due to physical characteristics among participants such as thick eyeglasses, oily skin, and thick eyelashes which prevented calibration. Three participants' data were partially removed from the study (data present in some but not all sets) due calibration loss mid-study (often caused by excessive head movements). This resulted in a net total of 20 (incomplete and complete data sets) consisting of 11 females and nine males. Each participant completed a consent form and reported vision or corrected vision sufficient to view the stimuli. Upon completion of the study, each participant was compensated \$10. The participant data used in each set are shown in the following three tables.

Table 1.

Participants Used in Set One

Kids-dog Human Figures	Family-statute Human Figures	Man-bottle Human Figures
Absent	Left	Right
1. Fe	1. Fe	1. Fe
2. Fl	2. Fl	2. Fl
3. Mb	3. Mb	3. Mb
4. Fh	4. Fh	4. Fh
5. Fk	5. Fk	5. Fk
6. Mg	6. Mg	6. Mg
7. Mk	7. Mk	7. Mk
8. Fa	8. Fa	8. Fa
9. Fd - removed	9. Fd	9. Fd - removed
10. Mh - removed	10. Mh - removed	10. Mh

Table 2.

Participants Used in Set Two

Man-bottle Human Figures	Kids-dog Human Figures	Family-statue Human Figures
Absent	Right	Right
1. Fb	1. Fb	1. Fb
2. Ff	2. Ff	2. Ff
3. Ma	3. Ma	3. Ma
4. Mf	4. Mf	4. Mf
5. Mi	5. Mi-removed	5. Mi

Table 3.

Participants Used in Set Three

Family-statue Human Figures	Man-bottle Human Figures	Kids-dog Human Figures
Absent	Left	Left
1. Fm	1. Fm	1. Fm
2. Fc	2. Fc	2. Fc
3. Fg	3. Fg	3. Fg - removed
4. Fj	4. Fj - removed	4. Fj - removed
5. Md	5. Md	5. Md
6. Mj	6. Mj	6. Mj

In set one, four of 30 trials were removed, resulting in an exclusion loss of 13%. In set two, one of 15 trials was removed, resulting in an exclusion loss of 7%. In set three, three of 18 trials were removed, resulting in an exclusion loss of 17%. While it is difficult to understand what an "expected" or "acceptable" loss percentage for a study might be (due to it not commonly being reported), the loss percentages in this study are comparable to findings of Crouzet, Kirchner and Thorpe (2010) who reported an exclusion loss of 746/3200 trials, or 23% for one of their experiments.

Materials

The three primary visual image stimuli used in this study were selected from a set of images studied by Wilkinson and Light (2011).

Criteria for Selecting Images. Several initial criteria had to be met in order for an image to be included in the study. First, the image had to feature humans as a prominent element. Second, the human figures featured in the images were required to strongly attract visual

attention, as determined by analysis of the fixation data collected by Wilkinson and Light (2011). The fixation data were examined between a period of five seconds for strong visual attention on the human figures. This particular timeframe was selected because it contains the earliest saccades which signify the areas of the image with the highest visual salience (Itti & Koch, 2001). The first two criteria were selected to address the research question addressing human saliency. The first criterion was useful as preliminary tool to identify images that could be used while the second criterion offered quantitative validity about the visual attractiveness of the images' human figures.

A third criterion was that images contain other elements that might be expected to compete for visual attention with the human figure. Photographs that contained too few visual elements (such as a portrait-style shot of a human figure standing in front of a blank wall) were eliminated. This criterion was important to find images that retained sufficient visual appeal even in the absence of human figures and also provided possible competitors for visual attention. The fourth criterion was that the researcher had to be able to adequately and convincingly digitally remove the human figures in Adobe Photoshop 7. This final criterion was especially significant because it identified images for which the spaces occupied by human figures could be made to appear to be seamless continuations of the image's background.

Experimental Conditions

The three conditions of the stimuli of human figures located on the left, human figures located on the right, and human figures absent yielded a total of nine unique stimuli. These nine images were arranged into three different sets so that each participant never saw more than one manipulation of an image. The following figure shows the order of the sets in rows in which participants viewed the stimuli. The top row of Figure 1 shows the three images that were shown

in both sets one and four, the middle row of Figure 1 shows the three images that were shown in set two, and the bottom row of Figure 1 shows the three images that were shown in set three.



Figure 1. Stimuli images listed in sets.

The arrangement of the images within the viewing sets is not totally randomized in part due to a re-categorization of the images done after participant testing. Initially, all images were referred to as "original", as in being the original orientation of the used in Wilkinson and Light (2011), "flipped", representing a mirror-image of the original orientation, and "absent", representing the condition with absent human figures. Thus, according to the original labels, each set follows a pattern of "absent, original, and flipped." For each set, participants viewed all three images in the same order, consistent with the method of Wilkinson and Light (2011). The images are described as follows (a) *Kids-dog*, which shows two children and a dog standing centrally in a scene that features a garden, a flower bush and a fence in the background; (b) *Family-statue*, which features a family consisting of a man and two children posed at the base of a statue with

five large human figures on top; (c) *Man-bottle*, which shows a man throwing a bottle into the air with and fountain in the background. Larger versions of all conditions of *Kids-dog*, *Family-statue*, and *Man-bottle* can be found in Appendices A: Figures A1-A3, B: Figures B1-B3 and C: Figures C1-C3, respectively.

Coding of Selected Images. To allow for analysis with the ISCAN technology, the images were coded by visual element. Coding involved using a selection tool to outline relevant objects in the image. This step enabled the ISCAN software to compute data on each "area" or element visually attended. Overall, this coding primarily identified elements found to be visually salient as indicated by the findings of other researchers (Niekamp, 1981; Vincent, Baddeley, Correani, Troscianko & Leonards, 2009; Chen & Zelinsky, 2006) which consisted of humans, color-contrasting elements, and elements implying motion (such as the fountain in the stimuli *Man-bottle*). Some elements were collapsed into the larger elements consisting of the backgrounds following the findings of Wilkinson and Light (2011) because they did not receive significant amounts of visual attention within the five second time period examined in that prior study.

Each of the images containing human figures were hand-coded individually, but for the absent human figures images, the coding scheme was copied from one of the other images and superimposed without having to be re-drawn. (See Appendices D: Figures D1-D3, E: Figures E1-E3 and F: Figures F1-F3.) Because the human figures in the left and right images were drawn separately, the coding schemes for each were analyzed for element size. The reason for the individual hand-coding was that while the coding scheme could be superimposed over a new image, it could not be horizontally manipulated to fit an image in the opposite (left/right) orientation. As a result, the differences between hand-coded orientations introduced a minor

limitation to data in terms of the unequal representation of the data from the left to right orientation.

After coding elements in each image, the ISCAN program reported the number of pixels occupied by each element and compared it to the total number of pixels in each image resulting in a percentage of the image occupied. For the left and right orientation conditions, the human figures were identified by coding a zone surrounding them. In the image without human figures, this same coding scheme was superimposed to identify the area the human figures would have occupied. The importance of maintaining space was to understand if there was anything about the area where the human figures had been (such as composition) that attracted visual attention. The criteria and element coding schemes per stimulus are as follows.

Kids-dog. The first image selected from the original image set of Wilkinson and Light (2011) features two children standing together in a garden as the human figures. As shown in Figure A2 (Appendix A: Figures A1-A3), *Kids-dog* also features a wooden fence located on the opposite side of the image from the human figures as well as a pink-and-red flower bush between the human figures and the fence. The background features a garden and a dark tree line and mountains are visible beneath an overcast sky. In addition to the initial selection criteria, *Kids-dog* was also selected because of other specific visual elements. For example, the fence appears on the opposite side of the human figures. This visual opposition provided an opportunity to study the possibility of a left-right bias between the fence and the human figures. The pink-and-red flower bush also fulfills a criterion of providing visual contrast because it contrasts in color to the dark green background. In addition, it provides possible competition for visual attention to the human figures because it is of a similar size and color as the human figures.

Table 1 shows the space occupied by the elements coded for this image. In addition to the individual elements noted, the background was split centrally down the image with the background featuring the human figures described as "HuBG" and the background opposite the human figures described as "AbBG".

Table 4.

Percent of Image Occupied for Kids-dog

Conditions	HuSub	Leaf	Flower	FlBush	Fence	HuBG	AbBG
HuSub Left	5%	1.1%	.7%	2.3%	2.5%	38.3%	45.6%
HuSub Right	4.2%	1.1%	.6%	1.8%	2.4%	36.2%	52.5%
and Absent							

Family-statue. The second image selected featured two children and a man posed at the base of a statue as human figures. As Figure B2 (Appendix B: Figures B1-B3) illustrates, this image also had "pseudo human figures" represented by three figures on the top of the statue and a medallion featuring the profile of a man's face. In the foreground at the base of the statue, there are two large red-and-white flower bushes on each side along with a red stop sign and the back of a car next to the larger flower bush. The background of this image has fewer visual elements and features trees of varying heights with an overcast sky. This image was selected because elements such as the statue and medallion might be possible competitors for visual attention because of their similarity to the actual human figures and the size differences between the pseudo and actual human figures. Also, like *Kids-dog*, *Family-statue* featured elements that provided visual contrast in terms of color. The red stop sign, red-and-white flower bushes and the illuminated taillights of the car all provide a sharp visual contrast to the mostly grey and green foreground and background colors. Although this image is more symmetrical than the

other two, it does still feature obvious differences in the position of the flower bushes and stop sign when manipulated by flipping the image across the vertical axis. Because of the centered location of the human figures, this image was classified as "HuSub Left" when the young girl in the pink shirt faced the left side of the screen and "HuSub Right" when she faced the right side of the screen.

In addition to human figures and the pseudo-human figures (statue and medallion), several elements of the foreground and background were also coded. While the car and stop sign were a part of this coding scheme, the red-and-white flowers were removed as individual elements due to the lack of visual attention toward them. Instead, they are coded as part of the overall background. Half of the background was coded as "SignBG" which represents the side of the image to the back of the girl in the pink shirt and featuring the stop sign and the other half was coded as "AbBG" which represents the side of the image to the front of the girl. Table 2 shows the comparison between the elements coded across the stimuli sets.

Table 5.

Percent of Image Occupied for Family-statue

Conditions	HuSub	Medallion	StopSign	StatueTop	SignBG	Car	AbBG
HuSub Left	3.9%	1%	.7%	16.6%	39%	.9%	34.6%
and Absent							
HuSub Right	3.9%	1%	.7%	17.1%	38.4%	1%	33.7%

Man-bottle. The third image selected from the image set featured a man in a plaza throwing a water bottle into the air. As Figure C2 (Appendix C:Figures C1-C3) illustrates, a large fountain is featured in the background and to either side of it are people which, due to their size and background location, are generally ignored during the time frame examined by this study (Wilkinson & Light, 2010). Behind the fountain are several buildings and trees which

appear under a bright blue sky with puffy clouds. Of particular interest in this image was its ability to fulfill the criterion of having implied motion. Implied motion is best described as any element of a still-frame photograph that was in motion when the picture was taken. In contrast, the opposite of implied motion would be any element that was in a stationary position when the picture was taken. To compare, *Man-bottle* is the only one of the three stimuli to feature implied-motion as *Kids-dog* and *Family-statue* are both stationary images. The two elements of this stimulus which imply motion are the human figure due to the interaction of his outstretched arm with the bottle which is seen in the air above him and the water fountain for the movement of the water.

Of all the stimuli, *Man-bottle* had the simplest coding scheme as it only accounted for the human figure (which was paired with the bottle due to their interaction), fountain, and the background which was split vertically and described by the presence or absence of the human figure. Unlike the previous stimuli, the coding for this image required that two unconnected areas of the background featuring the human figure be combined to create a total net element for the background. The following table compares the percentage of the image occupied by the elements across manipulations.

Table 6.

Percent of Image Occupied for Man-bottle

Conditions	HuSub*	Fountain	AbBG	HuBgSM**	HuBgBG**	HuBGNet
HuSub Left,	17.8%	16.3%	40.8%	3.2%	17.9%	21.1%
Absent						
HuSub Right	16.5%	15%	42.8%	3.9%	18.3%	22.2%

*HuSub was coded with the bottle because of the implied motion between the two elements.

**HuBgSM and HuBgBG were two elements that were combined to make HuBGNet, or the total of the HuBG.

Data Collection

Participant eye gaze data was collected through a setup of two computers and the ISCAN hardware and software. The participant sat beside the researcher at a desk with two computers. The computer facing the researcher featured the ISCAN software and allowed the researcher to calibrate the images, adjust for accuracy, and calibrate the infrared camera hardware for each participant. The infrared camera, an ISCAN©ETI-300 Binocular Free-head Eye laboratory system, was located in front of the participant and made no physical contact with the participant. A chinrest, as was used by Wilkinson and Light (2011), was offered to participants at their comfort. Pilot testing for a different study also using the ISCAN software revealed that the chinrest was not necessary for accurate data collection. Eye gaze was recorded by detecting light reflected off of the participants' corneas. Located behind the infrared camera was a 20" iMAC© which featured the stimuli at a resolution of 640x480 pixels. Each stimulus was presented for five seconds. The reason for the brief viewing time was to collect data on the participants' primary (and not consciously directed) views while allowing the researcher time to operate the ISCAN software which had to be coordinated and initiated for each image individually. While the short viewing time was necessary in order to capture visual attention to the elements of greatest interest, it also presents a limitation in applying these findings to longer view times.

To begin the experiment, participants were given simple verbal instructions to examine each image ("I'm going to show you an image for a few seconds and I just want you to look at the image.") This particular statement was used in order to create a "free viewing" context so as not to prejudice the participants' viewing pattern, as Yarbus (1967) found when specific verbal instructions altered his participants' scanpaths. This method is consistent with the language used by Wilkinson and Light (2011) in a similar study of visual attention. After viewing each image,

participants were presented with a blank screen and a small gray box centered toward the top of the screen. When the researcher saw the participants' eye gaze resting on the box, the researcher advanced to the next screen by clicking on it while simultaneously beginning the next trial in the ISCAN program.

Dependent Measures

Research data were in the form of quantitative data from the ISCAN software. Among other measures, ISCAN automatically outputs the measures of interest in this research, including latency to first view, the percentage of time spent on each element relative to the other elements, and the percentage of participants that viewed each element over time.

Because each participant viewed three images, it was found that the data from some participants were viable for some but not all images. In addition, an earlier coding scheme that was subsequently changed resulted in some sets having more participants than others. Instead of removing some participants to even the distribution of participants across sets, the numbers of participants in each cell was left uneven so that data would not be artificially excluded. To evaluate whether the uneven numbers might confound results, a split-half correlation was performed on the set of photographs on which data were obtained from the largest number of participants. The number of participants that viewed the absent human figures version of *Kids-dog*, the human figures left version of *Family-statue* and the human figure left version of *Man-bottle* were entered into a random number generator with each number representing a participants' data.

Table 7.

Number of Participants' Data Used Per Image

Image	Human Figures	Human Figures Left	Human Figures Right
-------	---------------	--------------------	---------------------

	Absent		
Kids-dog	8	4	4
Family-statue	6	9	5
Man-bottle	5	5	9

From this list of numbers, the total was split into two lists of random numbers. In the case of *Kids-dog*, each of the two lists contained four participants while in the case of *Family-statue* and *Man-bottle*, the first list contained five participants and the second list contained four. A Pearson correlation was then performed to analyze the degree of similarity or difference between the two randomized lists of participant data. Each of the split-half analyses was found to be highly correlated.

Table 8.

Split-half Analyses of Selected Image Conditions

	Kids-dog: Human figures absent	Family-statue: Human figures left	Man-bottle: Human figure left
Pearson correlation	.85	.85	.91

Correlations ranged from .85-.91. This suggests that the data were highly consistent irrespective of whether the larger full set or smaller subsets were included. For this reason, data were included for the larger full set during analysis.

Last, for the analysis of latency to first view of individual elements, artificially high numeric values were used to prevent the results from being deflated when calculating the statistical analyses due to the way the ISCAN software recorded data. Because the program recorded a time to first view of zero seconds for elements that were not viewed (implying that the elements were viewed instantaneously with no latency to first view rather than having never

been viewed), values of zero seconds were replaced with a value of five seconds which was the maximum viewing time per trial. Therefore, when mean latency times were calculated, elements that were not visually attended did not deflate the mean and inaccurately represent the latency to first view.

Results

I began by examining the fixation patterns when no human figures were present, then considered how the fixation changed when the humans were added either to the left-or-right-hand side of the image. To address the research question concerning the salience of human figures within a VSD, the average time to the first view of the human figures when oriented to the left and right and when absent was analyzed. To address the research question concerning the effects of the location of the human figures, participants' visual fixations to the human figures for each image were compared to the other elements across each condition. Finally, a ratio derived from the work of Fletcher-Watson et al. (2008) was created based on the size of each image's elements as a comparison of the amount of time which was spent viewing them versus the amount of time that would be expected based on the element's size. This last analysis allowed for the proportionality of the elements' size to be taken into consideration when compared to the amount of time spent viewing the elements. For example, an element that occupies 5% of an image would not be expected to attract the same amount of attention as an element that occupies 50% of an image. This ratio was especially important in understanding how human figures attracted attention because the amount of time participants' spent viewing them was unexpected given their size.

Kids-dog

When there were no human figures, the first element to capture attention on average was the right side of the background (mean latency to first view = .49 seconds). Views to specific elements, including the combined leaf/flower/bush element and fence, did not occur until .91 and 1.49 seconds, respectively. When the human figures were present and appeared on the left-hand side of the image, the human figure was the first element to attract attention, with a mean latency

of .56 seconds, followed by the background behind the human subjects (2.37 seconds latency). When the human figures appeared on the right-hand side of the image, the human figures were the first element to attract attention, with an average latency of .92 seconds, followed by the background (1.79 seconds). In summary, when the human figures were present, they attracted attention within one second of the image appearing, and were either the first or the second element to attract attention. Figure G1 in Appendix G: Figures G1-G3 shows these data. The dotted vertical lines indicate the standard deviation of the mean.

I also examined how much of the total viewing period was spent on each coded element, to evaluate how much attention was dedicated to each element. During the total viewing period, participants dwelled on the defined elements of interest for between 50-68% of the time (the remainder of the time consisted of time the participant spent blinking or moving between elements, that is, engaged in rapid saccades). Considering only the time spent dwelling on the defined elements, Figure G2 in Appendix G: Figures G1-G3 presents a stacked bar chart illustrating the distribution of time spent on each element. Not surprisingly, in the condition with the humans absent, the area where the humans would have been was attended only briefly (2% of the dwell time). The background not behind the human attracted the most attention (48%) followed by the leaf/flower/bush (21%) and the fence (16%). In contrast, when the human figures were present, they garnered more attention than any other element, and indeed, more than half the time spent on elements was spent on the humans (in both orientations).

The final analysis evaluated the ratio of each element's size and the amount of time participants spent viewing the element. The ratio provides information as to whether each element viewed was viewed more or less than would be expected based on the element's size. Elements with a ratio of less than one were considered to have been viewed less than was

expected based on their size while elements with a ratio of greater than one were considered to have been viewed more than was expected based on their size. The dotted vertical lines indicate the standard deviation calculated from each element's mean. Lines that remain above the ratio of one indicate that the data are considerably different from the ratio.

Figure G3 in Appendix G: Figures G1-G3 presents the ratios of time spent relative to space occupied under the three experimental conditions. In the absent condition, the fence and combined leaf/flower/bush elements received substantially more views than would be expected relative to their size. When the human figures were present, views toward these non-human elements decreased. However, the amount of time spent viewing the human figures was greater than would be predicted based on their sizes, at 4.55 (human figures on the left) and 9.52 (human figures on the right). The ratios for examining the humans are larger than the ratios for any other coded element, including for the elements viewed when no humans were present. The dotted vertical bars in this image represent the standard deviation of each element's mean.

Family-statue

With human figures absent, the first element to capture attention on average was the pseudo-human figure group on top of the statue (mean latency to first view = .28 seconds). Views to specific elements, including the medallion (representing another pseudo-human figure due to the man's profile) and the stop sign did not occur until 3.39 and 4.61 seconds on average, respectively. When the human figures were present on the left-hand side of the image, they were the second element to be viewed (mean latency = 1.45 seconds) after the pseudo-human statue top figures (mean latency = 1.28 seconds). When the human figure appeared on the right side of the image, they were the first element viewed at 1.11 seconds followed by the pseudo-human figures of the statue top (mean latency to first view = 1.83 seconds). Thus, when human figures

were present, they were viewed within 1.5 seconds of the image appearing (for both sides). When the human figures were not the first element viewed, they were the second after the centrally located and prominent pseudo-human statue. These data can be found in Figure H1 in Appendix H: Figures H1-H3.

In this image, the amount of the total viewing period that participants spent dwelling on the defined areas of interest ranged from 46-62% with the remainder of the time consisting of blinks, saccades, looks away, etc. The distribution of visual attention to each coded element during these dwells is illustrated in Figure H2 in Appendix H: Figures H1-H3. Again, it is no surprise that the area the human figures would have occupied in the condition without human figures was visually attended by none of the participants (0% of total viewing time). The pseudo-human figures of the statue top attracted the most attention (72%) followed by the pseudo-human profile on the medallion (13%) and the right side of the background (7% with stop sign). In contrast, when the human figures were present, they competed for attention with the statue top by decreasing the amount of attention the statue top received compared to the condition with no humans. When the human figures were on the left-hand side, they received nearly as much attention (35%) as the pseudo-human figures (43%). When the human figures appeared on the right, they were visually attended 27% of the time, but the pseudo-human figures were viewed approximately twice as much (50%). Overall, the pseudo-human figures received proportionally more time when the human figures were on the right while there was a much less significant difference in viewing when the human figures appeared on the left.

In addition, we found that several elements were viewed by participants considerably more than would be considered based on their size. Figure H3 in Appendix H: Figures H1-H3 presents these findings. In the condition with no human figures, the medallion received

substantially more views and the pseudo-human figures on the statue top received moderately more views than would be expected relative to their size. When the human figures were present, views to non-human elements (including pseudo-human figures) were found to greatly decrease while the amount of time spent viewing the human figures was substantially greater than would be predicted based on their sizes at 5.01 on the left and 4.36 the right side. Also interesting was the finding that the car received more views than would be expected relative to its size when human figures appeared on the left (2.47), and more views than would be expected when human figures appeared on the right (5.60).

Man-bottle

For the final image, when the human figure was absent, the first element to capture attention was the fountain (mean latency time = 1.49 seconds). Views to other elements such as the right side of the background and the area currently unoccupied by the human figure lagged considerably (2.38 and 2.49 seconds, respectively). When the human figure was present and appeared on the left-hand side of the image, the human figure was the first element to attract attention, with a mean latency of .52 seconds, followed by the background opposite the human figure (1.97 seconds). When the human figure appeared on the right side of the image, the human figure was again the first element to attract attention (.57 seconds) followed by the fountain (1.72 seconds latency). In summary, when the human figure was present, it attracted attention shortly after the first half second of the image appearing and was always the first element to attract attention. These results are shown in Figure I1 in Appendix I: Figures I1-I3.

For the analysis of how much of the total viewing time was spent on each coded element, between 50-68% of the total viewing time included dwells on the areas of interest. Viewing of the area that would have been occupied by the human figure lasted 10% of the total viewing

time, as seen in Figure I2 in Appendix I: Figures I1-I3. The fountain received the most attention (51% of the dwell time) followed by the left side of the background (31%) and the right side of the background (7%). When the human figure was present, it attracted more attention than the remaining elements in either orientation (30% and 32% for the left and right orientations, respectively).

The final ratio analysis performed on this image showed that in the absent condition, the fountain received more views than would be expected relative to its size. As Figure I3 illustrates (Appendix I: Figures I1-I3), when the human figure was present, views to the fountain as well as other elements decreased. In addition, the amount of time spent viewing the human figure is slightly above what would be predicted based on its sizes at 1.46 (human on the left) and 1.68 (human on the right). When the human figure was present, the ratio for examining it was still larger than the ratios for the remaining coded elements.

Discussion

Overall, humans attracted initial visual attention in the first five seconds of viewing a scene in a photograph when they were present, regardless of their orientation and the other coded elements the image contained. Consistent with recent studies from other laboratories (Crouzet et al., 2010; Fletcher-Watson et al., 2008; Humphrey & Underwood, 2010), my analysis of the participants' fixations supports humans as strongly salient visual features of a visual scene.

Another general trend observed was that for each of the images, neither latency to first view nor average viewing time supported the left-orientation bias such as that I predicted would occur on the basis of findings by Nummenmaa, Hyönä and Calvo (2009). While it is not understood why this study failed to support their findings, it is possible that the varying size and prominence of the human figures may have influenced viewing patterns or reduced any orientation bias. This discussion will systematically review each of the images separately, beginning with *Kids-dog*.

Kids-dog

For this image, the findings show several patterns in participants' views to the human figures. First, when human figures were present, they were either the first or second element viewed and were visually attended within the first second of viewing. This speaks to the strong attraction to humans in scenes, consistent with data from Wilkinson and Light (2011).

The conditions with the humans on the left and right had very similar viewing patterns because the error was much smaller for the human figures than for the area occupied by the human figures in the absent condition. The remarkable way in which human figures were viewed so similarly by participants while the other elements were viewed so differently further supports the need for more research in the area of visual scene displays in order to better understand the mechanisms that affect visual attention.

With the ratio analysis, it was found that the human figures received much more visual attention than would be expected based on their size. The ratio also revealed what appeared to be a shift in attention toward the non-human elements in the absent condition because they attracted attention similar to that seen with the human figures. When human figures were present on the right, visual attention to the other elements decreased with only the fence and leaf/flower/bush receiving more attention than would be expected based on their small size. When human figures were present on the left, the human figures, leaf/flower/bush and fence all attracted attention similarly. However, although the leaf/flower/bush and fence were viewed more than would be expected based on their size, the human figure had the highest ratio indicating that it was viewed considerably more than would be expected. This finding tentatively explores the relationship between human figures and other elements in visual scenes.

Finding that with this image, the diminutive size of human figures can attract attention much greater than would be expected based on their size may have implications for future research concerning VSDs. For one, it would be worthwhile to explore this interaction in more depth to discover if there is a reliable relationship between the size of human figures and visual attention and if there is a point at which the size of a human figure no longer attracts more visual attention than would be expected based on its size. Research such as this could help to provide guidelines for VSD programming. Perhaps the size of human figures can be manipulated in order to better attract visual attention, especially for individuals for whom the maintenance of visual attention can be challenging. While the interaction between the human figures and non-human elements in this image are not well understood, it is important to note that the increase of visual attention to human figures is associated with a decrease of visual attention to non-human elements.

Family-statue

Despite the presence of many other potential distracters in this photograph (the car, flag, flowers, and so forth), the majority of visual attention focused on the actual (live) human figures and the pseudo-human statue figures. During the five second viewing time, it is likely that the participants were too preoccupied with the foreground elements such as the statue to visually explore the right and left background which primarily feature the trees and sky. This finding reinforces the finding that humans capture initial attention when present and underscore their centrality to visual attention. Understanding this characteristic of viewing humans may affect construction of VSDs by encouraging the use of more naturalistic images with context-rich backgrounds. The background will allow viewers to infer more about the scene without the viewer's visual attention being distracted by or lost on other elements. For example, symbol representations may be better realized by having human figures appear before a background such as a school, playground, or home setting rather than having blank white backgrounds. Furthermore, the backgrounds could be uploaded images relevant to that particular individual which would provide even more context and personalize the image.

The fact that the orientation of the human figures (and pseudo-human figures) caused differences in visual attention was surprising especially due to their near-unchanging central location. The statue top, for example, was a considerably strong attractor of visual attention but was viewed faster (and in fact, before the human figure) when the human figures were facing the left side of the image. When the human figures faced the right side of the image, the human figures were viewed first and the statue top lagged by approximately .7 seconds. This finding warrants more systematic research with greater control over the images, to see if it will be replicable. Another interesting finding that warrants more research is the apparent divide

between the three-dimensional statue top and the simplified depiction of a human, such as the medallion, which featured the side profile of a human. Even though the medallion was viewed in the absent condition, it was all but ignored when the human figures were present. In contrast, the statue top attracted visual attention similar to the human figures across all three conditions.

While it was hypothesized that the pseudo-human figures would attract visual attention similar to that of the true human figures, it was remarkable to find how closely their viewing patterns matched those of the human figures. In fact, the statue top in the absent human figures condition was perceived to attract visual attention nearly identically to how a true human figure might. One possible explanation is its proximity to the top center of the image (near a grey box used to standardize participant visual attention between images) and participants' visual attention before the image appeared. However, while this explanation may provide insight into viewing of the statue top in the absent human figure condition, it does not explain why first view of the human figures and statue top were reversed in the conditions with the humans on the left and right. Clearly, future research on issues of animacy is warranted.

Even in spite of these differences, the true human figures were found to have a systematic effect on visual attention when their total size relative to the size of the image is considered. With human figures occupying only 3.9% and the statue top occupying between 16.6-17.1% (varying with coding) of the image across all conditions, it is important to note that the human figures competed for visual attention with elements that were approximately four times their size. This particular finding further supports the importance and strength of human figures as attractors of visual attention.

Similar equality in viewing the human figures when on the left and right was also observed in the analysis of the elements in *Family-statue* image. The presence of human and

pseudo-human figures in this image led to the unexpected finding that both elements were viewed similarly. For example, the statue top, which represents three pseudo-human figures that occupy a large percentage of space within the image were viewed similarly to the actual human figures. In fact, viewing toward the pseudo-human figures was not found to decrease as significantly as, for example, the non-human elements in the *Kids-dog* image. Also interesting to note was that the medallion, which represented a relatively small and less realistic pseudo-human figure, was viewed in a pattern consistent with the non-human elements in *Kids-dog* in that it received very strong visual attention only when the human figures were absent. However, the visual attention it received in the absent condition was greater than the visual attention given to the human figures based on their size. While this particular finding would necessitate further research in order to better understand this possible effect, the use of human figures and VSDs in AAC could lead to important developments in facilitating individuals' ease of interaction with their AAC devices by encouraging visual attention.

Examining the elements as ratios comparing their size and the amount of time spent viewing them allowed for much different conclusions to be drawn about participants' visual attention. While the pseudo-human figure statue top was strongly visually attended throughout the three conditions, the ratio showed that the time spent viewing this element may have been a result of the element's size rather than it being more appealing to participants than the true human figures. Additionally, it is also likely that focusing of participants' eye gaze on a grey box located centrally and above the top frame of the image between trials resulted in the statue top being the closest element for viewing which caused it to receive more views as well. It is unknown what affect (if any) would be caused by positioning the participants' focus box

differently, such as to the left or right. Further research could address the effect of a "visual starting point" on visual attention and how participants view the following stimulus.

Man-bottle

This final image presented participants with the largest and most prominent human figure out of all the images. In addition, it also included "implied motion" due to the movement of the water spray in the fountain as well as the human figure throwing a bottle which is suspended midair. The fountain, which was also a large and prominent element, was a strong competitor for visual attention with the human figure. This image presents an "attention shift" seen when participants' visual attention changed dramatically when the human figure is absent compared to when it is present. In the absence of the human figure, the fountain attracted and maintained the majority of participants' visual attention. This resulted in minimal views toward other coded elements in the image. However, when the human figure was present, participants viewed it more quickly than fountain in the absent condition. Furthermore, the fountain continued to lag behind views to the human figure on both the left and right. It was also interesting to note that more visual attention to the background behind the human figure was noted in the conditions containing the human figure.

The overall size of the elements in this image may have resulted in a decrease in the differences between left and right human figure placement that was observed in the previous two images. With the exception of views to the background behind the human figure, participant viewing habits were similar for the left and right conditions. It is unknown what exactly caused this similarity in viewing, whether it was the implied motion of both the human figure and the fountain, or the fact that the elements occupied such a large amount of space

within the image so as to make their placement within the scene irrelevant in terms of attracting visual attention.

Perhaps most compelling is the ratio by which the elements attracted attention relative to their size. Overall, the ratios were far lower for this image than for either of the two other images. The greatest ratio, which was slightly over two, was that of views to the statue in the absent condition. In both the left and right condition, the human figure was the only element viewed more than would be expected based on size. Another possibility of the low ratios could be that they were negatively affected by the implied motion. While *Kids-dog* and *Family-statue* are considered "static" in that they do not display action, but rather posed and stationary images, this image shows implied motion in both the fountain and the human figure. This interaction between implied motion, size, and the human figure is yet another area for further research. While it may be that there is a minimum size for human figures to attract attention, there may also be a maximum size by which human figures lose (or lessen) their ability to be strong attractors of visual attention.

Summary and Conclusion

The findings of this study in relation to the original hypotheses are as follows:

1) As hypothesized, human figures were found to strongly attract visual attention as measured by latency to first view when competing with non-human elements as well as total viewing time relative to other elements and ratio of views received relative to size. An unexpected part to this hypothesis was the finding that pseudo-human elements received visual attention similar to that of true human elements.

2) Contrary to the original hypothesis, visual attention to objects within the image were not found to differ based on the image's orientation. Human figures especially were viewed similarly (in terms of average viewing time and latency to first view) regardless of orientation within the scene. There were some individual variances in early views to non-human elements, but it is believed that these are more a result of individual variation due to the relatively small number of participants.

3) Consistent with the original hypothesis, human figures were found to affect viewing patterns. When human figures were absent, the average latency to first view of the other elements was different than when the human figures were present. In other words, the presence of human figures caused most of the elements to be viewed in a different order. Furthermore, in some conditions, elements that were viewed when human figures were absent were ignored when the human figures were present. This viewing phenomenon was unexpected and requires further research to better understand why visual attention to human figures disrupts visual attention to other non-human elements.

Applying Findings to the Field of Communication Sciences and Disorders

These findings provide new information that may help SLPs and other related professionals to create more effective Visual Scene Displays. By understanding that human figures in scenes are more visually salient than scenes without human figures, SLPs may be able to develop VSDs that strongly attract the visual attention of both the users of AAC and their communicative partners. It may be that these human figures can be placed freely within the scene, without worry that placing them in a specific orientation will lessen their attraction as visual elements.

In addition, these visual scenes can include features perceived to compete with human figures for visual attention without actually doing so. It is hoped that increasing the use of different photographic elements such as implied motion as well as contrasting colors may improve the quality of VSDs by offering more as opposed to less. Even considering the visual complexity of the stimuli studied, it is important to understand that the human figures were the strongest attractors (in terms of latency to first view and average viewing time) of all the elements competing for visual attention.

Limitations and Future Research

While this research made an important contribution to further understanding visual attention to VSDs, it is also important to critique the study's limitations in order to improve future research. The primary limitations of this study involve numbers, in that the systematic replication examined only three base pictures (although there were a total of nine including manipulations) and that the number of participants were spread unevenly throughout the groups. The relatively small number of image stimuli used was a limitation because while thorough information is available with these stimuli, the findings are unable to be generalized to other

images. Additionally, because these images were not controlled for variables such as left/right distance from the center or size of the human subjects across the three stimuli, it is unknown whether the findings could be affected by these variables.

The uneven distribution of participants could also be avoided by using image sets that do not involve repetition. While the repetition of one set (weighting it more heavily than the other two sets) is a limitation of the study due to the late revision of the original coding scheme, it was important that the data collected was retained as it still displays the trends found with the image sets with half the number of participants. Future studies will benefit from the unevenly weighted sets which, in hindsight, were an unnecessary attempt to balance participants across all sets. Lastly, the final limitation is one that is common throughout research and is the lack of participant diversity. Because all participants were students enrolled at The Pennsylvania State University, the findings are limited in their ability to be generalized across diverse populations of individuals. The obvious way for researchers to address this limitation is by sampling broader populations of individuals, something that is often limited by geographic location (such as the isolated State College, PA area), participant attendance during data collection sessions, and resources.

Furthermore, it is hypothesized that the data collected during the short view period used in this study (five seconds) may be relatable to longer viewing periods as the patterns of visual attention (specifically those to human figures) are similar to the findings of Yarbus (1967) from a study that used a three minute viewing time. By examining visual attention for only five seconds, data were collected on the most visually salient elements (specifically, the human figures) within the scene. However, this limits the ability of the data to be generalized to longer viewing times, multiple views of the same image, and multiple views of the same image over

time. Based on the similarity in viewing patterns to the longer Yarbus (1967) study, I hypothesize that the visual attention recorded during this brief initial viewing time is comparable to visual attention over a longer period of time and that a longer viewing period does not necessary imply a difference in visual attention. Conversely, further research is required in order to better understand visual attention during multiple trials of the same visual scene as well as multiple views over time (days, weeks, etc.).

Finally, this study was limited to the lack of scientific control because each of the base images was unique and not controlled for criteria such as size of the human figures or distance from the center of the image (to better define "left" and "right"). Despite these issues, study of naturalistic scenes is important in order to better understand what mechanisms affect visual attention. In this way, VSDs for use in AAC may be reverse-engineered. Unfortunately for the timing of this study, a similar study by Humphrey and Underwood (2010) was not available when the methodology and data collection for this study was developed although it is clearly an area of interest that will benefit from continued research.

It would also be interesting to see if there is a relation in visual attention to further broken-down depictions of the human figure such as a stick figure or line drawing rendering. This type of research could help professionals to better understand if individuals using alternative symbol systems to communicate perceive the "human" representations as truly human. Another area of possible research and growth would be the integration of pre-existing technologies (such as those available in cell phones and computers) into AAC systems. It is important that users of AAC have the same access to the same tools of connectivity and data-sharing so that communication is made easier and more accessible. It seems unacceptable that images can be quickly and easily uploaded to networking sites, applied to software that analyzes

them for faces, and labeled by individuals in less than a minute while the programming of AAC devices remains a difficult and time-consuming process completed by specialists. The disparity is made even greater considering that the technology has been invested in a leisure activity such as social networking, as opposed to the fundamental need to communicate.

These goals will likely require the collaboration of researchers from a variety of backgrounds and specialties. Thus, it is important that researchers specific to the field of communication sciences and disorders not only seek other professionals to assist in their research, but also to disseminate the results of their research so to maximize its potential benefits. Lastly, it is important that these studies include participants with diverse communication challenges in order to understand if visual processing is universal across all humans, or may vary based on specific disorders or even severities of specific disorders.

References

- American Speech-Language-Hearing Association. (2002). *Augmentative and Alternative Communication: Knowledge and Skills for Service Delivery* [Knowledge and Skills]. Retrieved February 13, 2011 from www.asha.org/policy
- Blackstone, S.W. (2004). Visual scene displays. *Augmentative Communication News*, 16, 1-16.
- Buswell, G.T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: The University of Chicago Press.
- Castelhano, M.S. & Henderson, J.M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology*, 62(1), 1-14.
- Chen, X. & Zelinsky, G.J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, 46, 4118-4133.
- Crouzet, S.M., Kirchner, H. & Thorpe, S.J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, 10(4), 1-17.
- Fletcher-Watson, S., Findlay, J.M., Leekam, S.R., & Benson, V. (2008). Rapid detection of person information in naturalistic scene. *Perception*, 37, 571-583.
- Humphrey, K. & Underwood, G. (2010). The potency of people in fixations: Evidence from sequences of eye fixations. *Journal of Vision*, 10(10), 1-10.
- Itti, L. & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203.
- Light, J.C., Drager, K.D.R., & Nemser, J.G. (2004). Enhancing the appeal of AAC technologies for young children: Lessons from the toy manufacturers. *Augmentative and Alternative Communication*, 20(3), 137-149.

- Light, J.C., Page, R., Curran, J. & Pitkin, L. (2007). Children's ideas for the design of AAC assistive technologies for young children with complex communication needs. *Augmentative and Alternative Communication*, 23(4), 274-287.
- Niekamp, W. (1981). An exploratory investigation into factors affecting visual balance. *Educational Technology Research and Development*, 20(1), 37-48.
- Noton, D. & Stark, L. (1970). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11, 929-942.
- Nummenmaa, L., Hyönä, J., & Calvo, M.G. (2009). Emotional scene content drives the saccade generation system reflexively. *Journal of Experimental Psychology*, 35(2), 305-323.
- Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, 17(6), 812-834.
- Vincent, B.T., Baddeley, R., Correani, A., Troscianko, T., & Leonards, U. (2009). Do we look at lights? Using mixture modeling to distinguish between low- and high-level factors in natural image viewing. *Visual Cognition* 17(6), 856-879.
- Wilkinson, K.M. & Light, J. (2011). *Visual attention to human figures in photographs: Considerations for the design of aided AAC displays*. Manuscript submitted for publication.
- Yarbus, A.L. (1967). *Eye movements and vision*. New York: Plenum Press.

Appendix A



Figure A1. Kids-dog with human figures absent.

Appendix A



Figure A2. Kids-dog with human figures on left.

Appendix A



Figure A3. Kids-dog with human figures on right.

Appendix B



Figure B1. Family-statue with human figures absent.

Appendix B



Figure B2. Family-statue with human figures on left.

Appendix B



Figure B3. Family-statue with human figures on right.

Appendix C



Figure C1. Man-bottle with human figure absent.

Appendix C



Figure C2. Man-bottle with human figure on left.

Appendix C



Figure C3. Man-bottle with human figure on right.

Appendix D

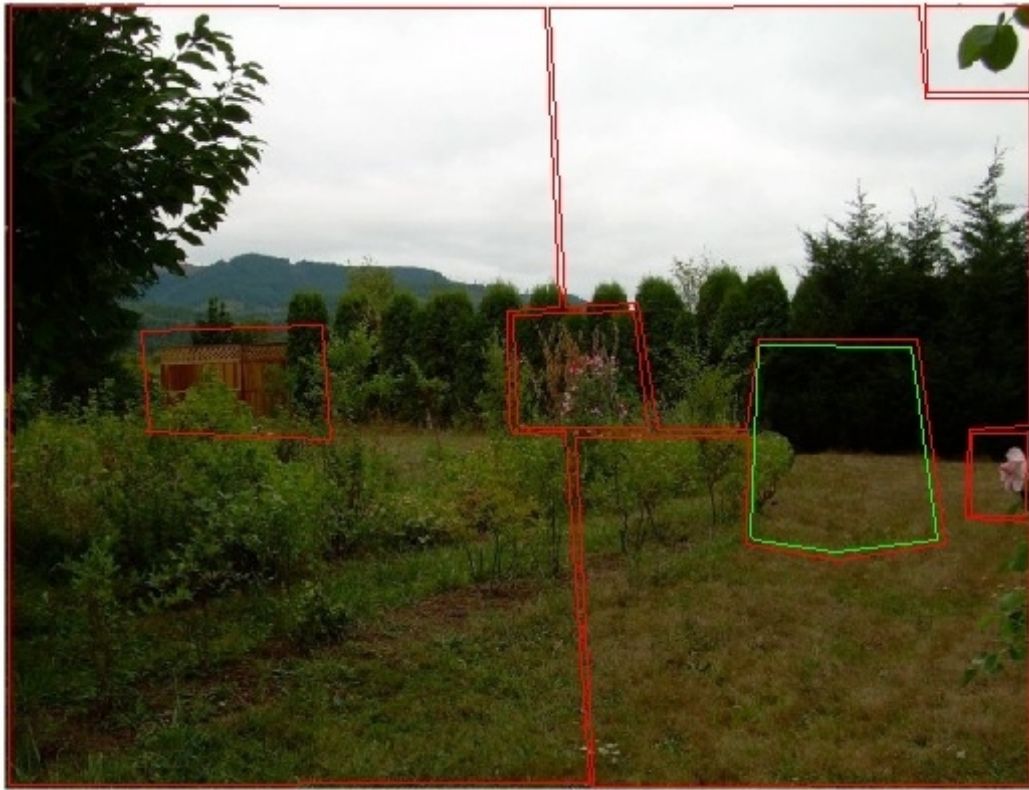


Figure D1. Kids-dog: Element coding with human figures absent.

Appendix D

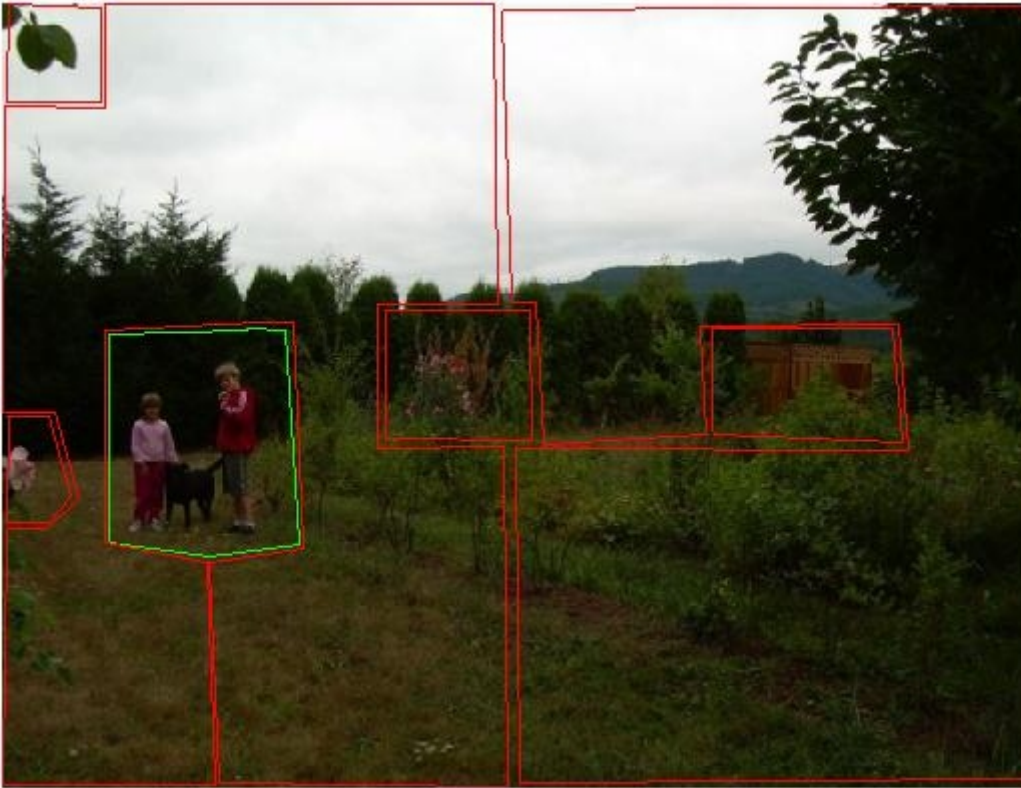


Figure D2. Kids-dog: Element coding with human figures on left.

Appendix D

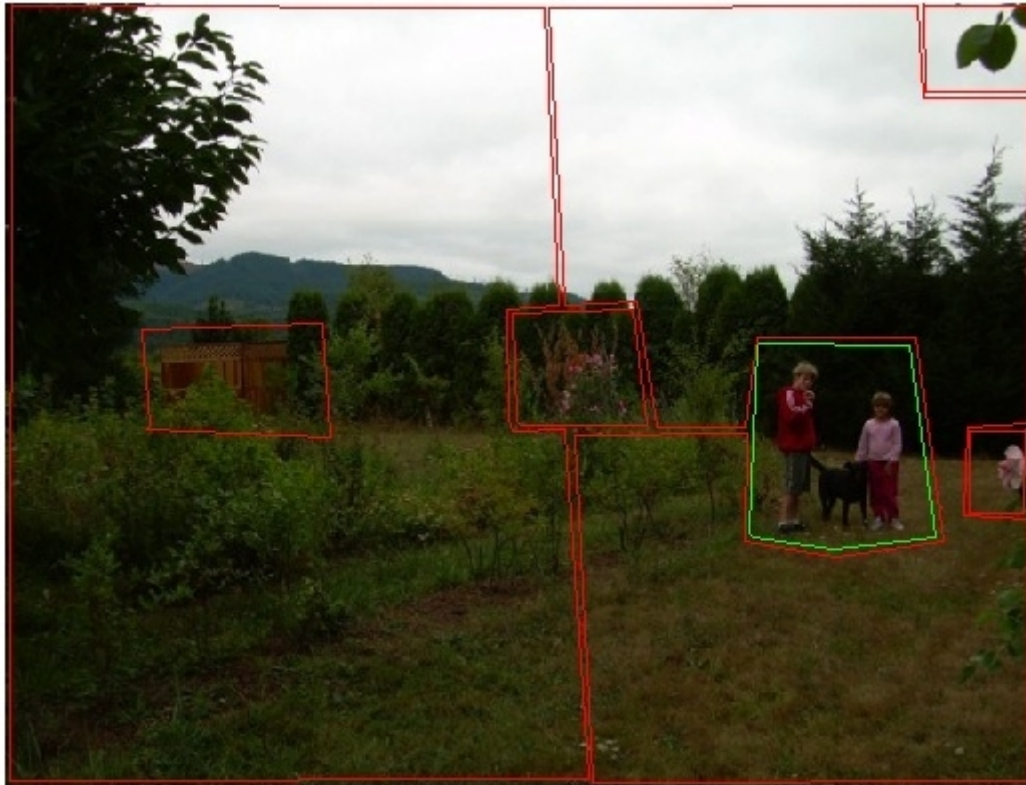


Figure D3. Kids-dog: Element coding with human figures on right.

Appendix E



Figure E1. Family-statue: Element coding with human figures absent.

Appendix E

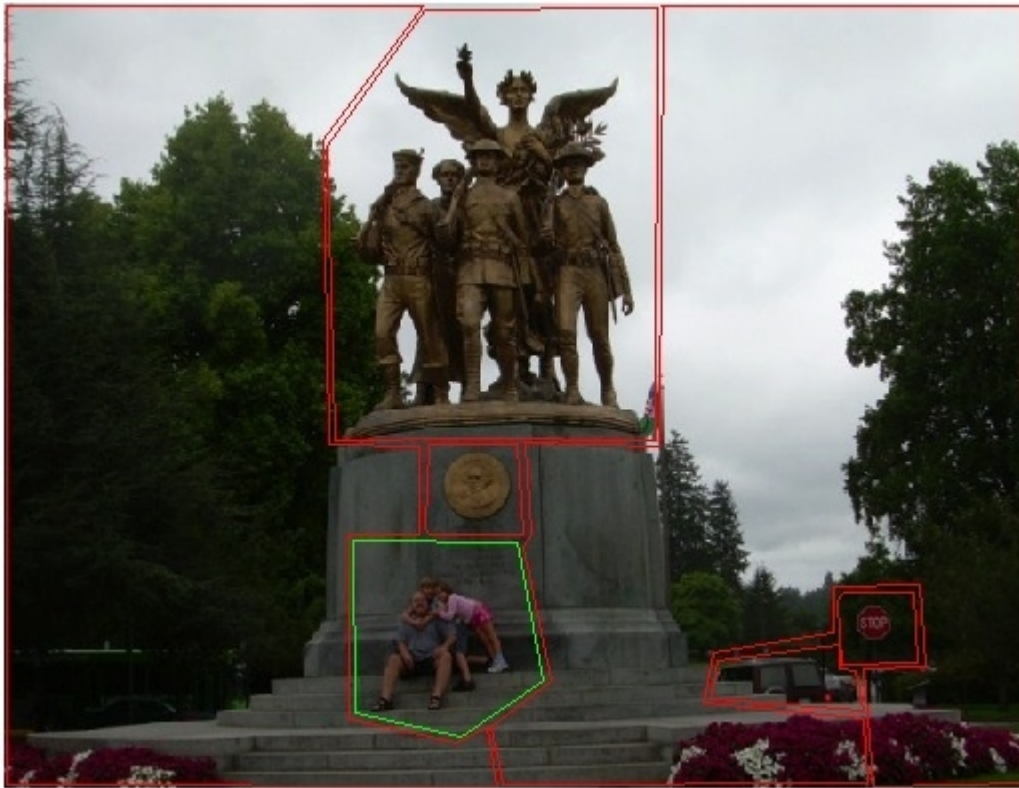


Figure E2. Family-statue: Element coding with human figures on left.

Appendix E



Figure E3. Family-statue: Element coding with human figures on right.

Appendix F

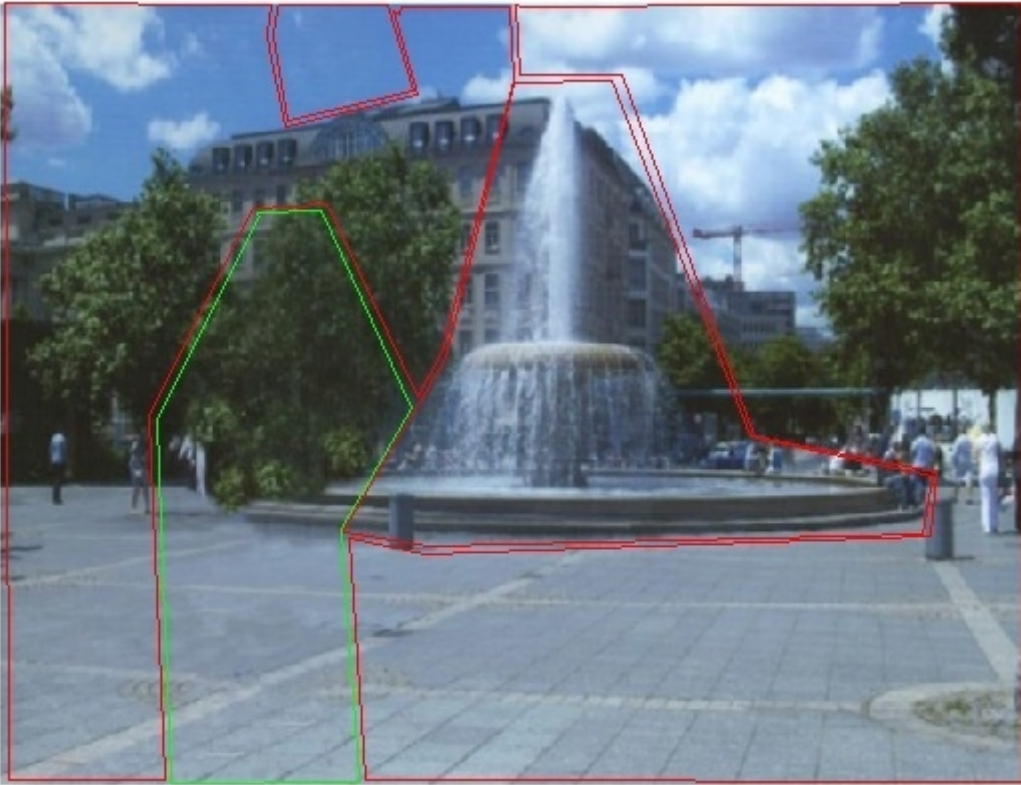


Figure F1. Man-bottle: Element coding with human figure absent.

Appendix F

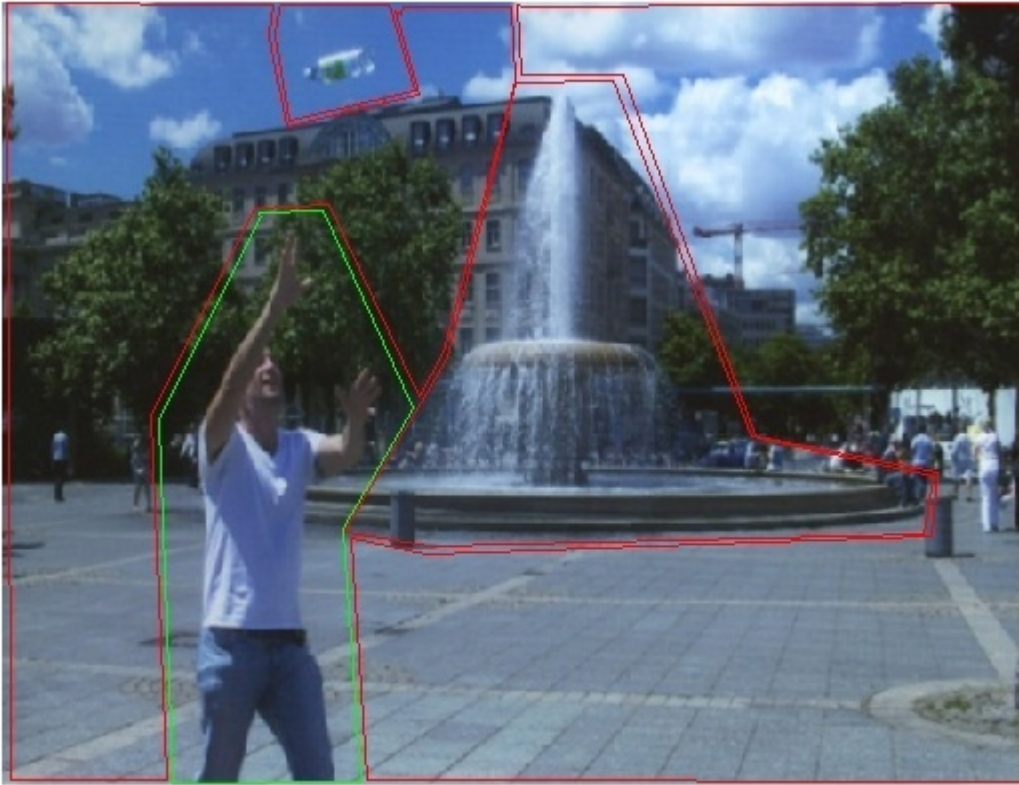


Figure F2. Man-bottle: Element coding with human figure on left.

Appendix F

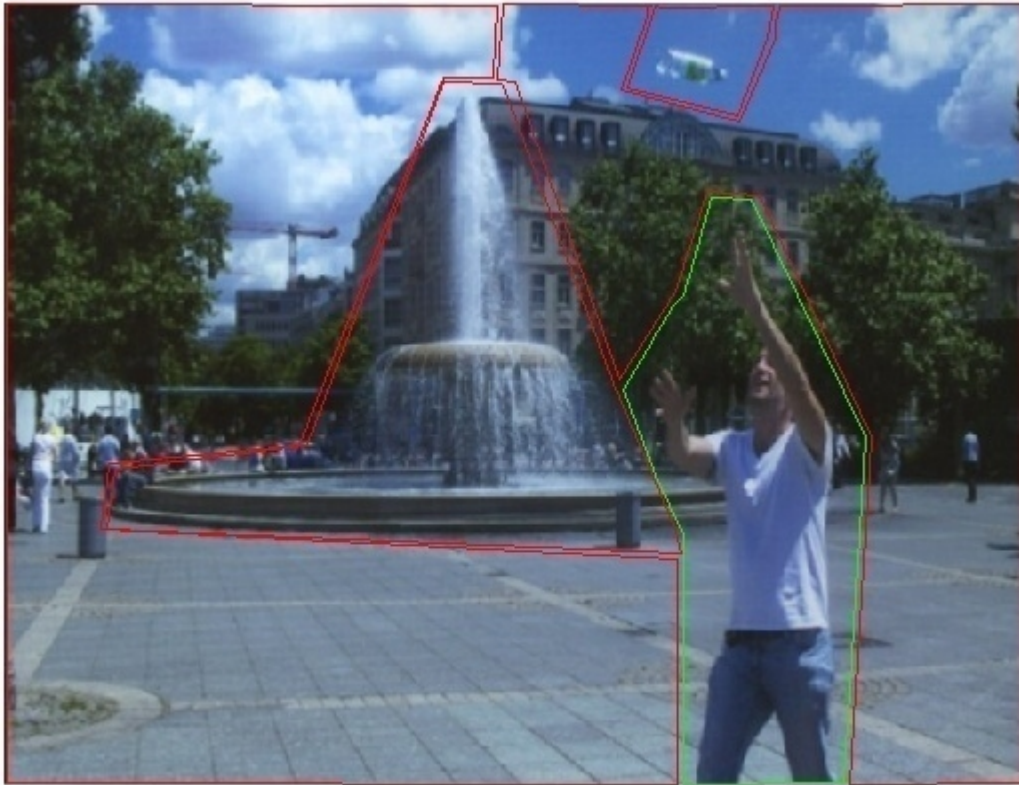


Figure F3. Man-bottle: Element coding with human figure on right.

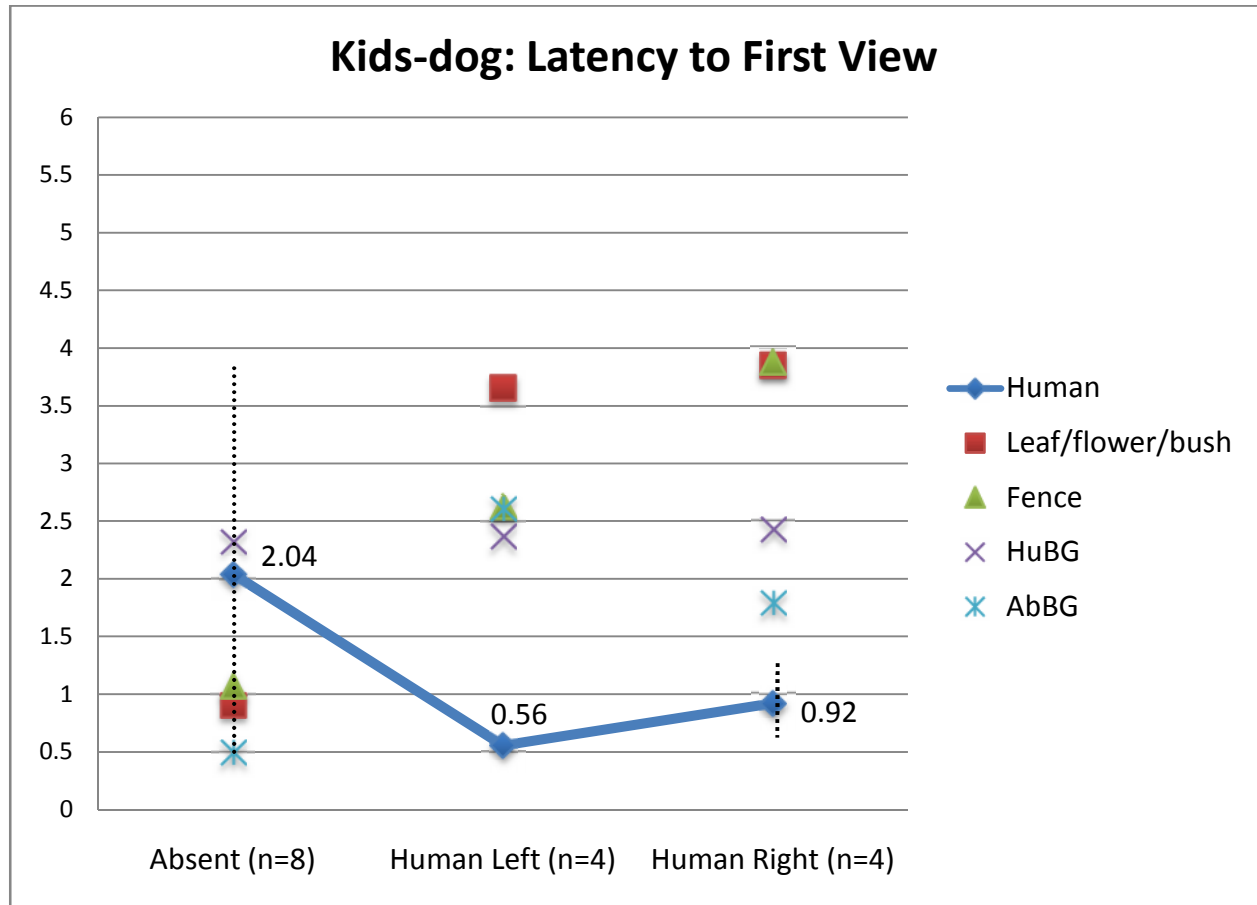


Figure G1. Kids-dog: Latency to first view.

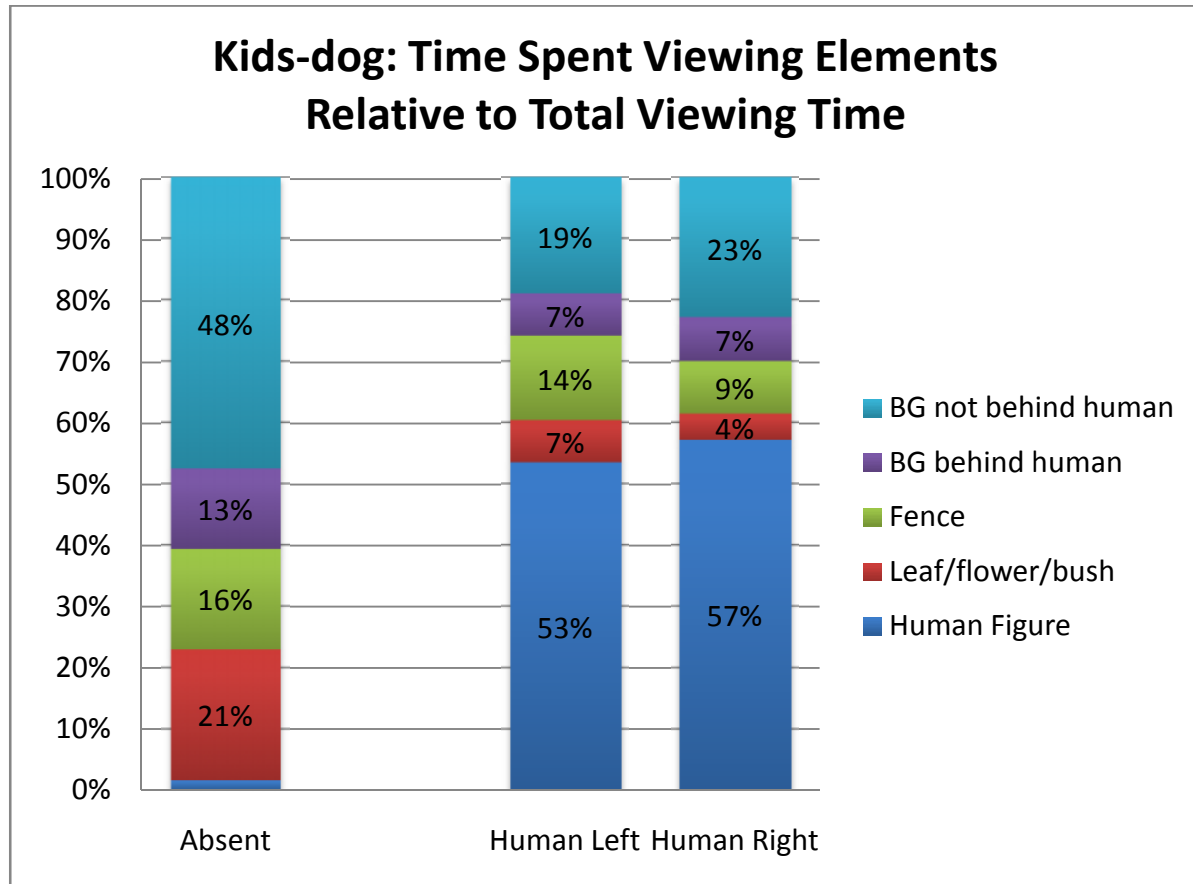


Figure G2. Kids-dog: Time spent viewing elements relative to total viewing time.

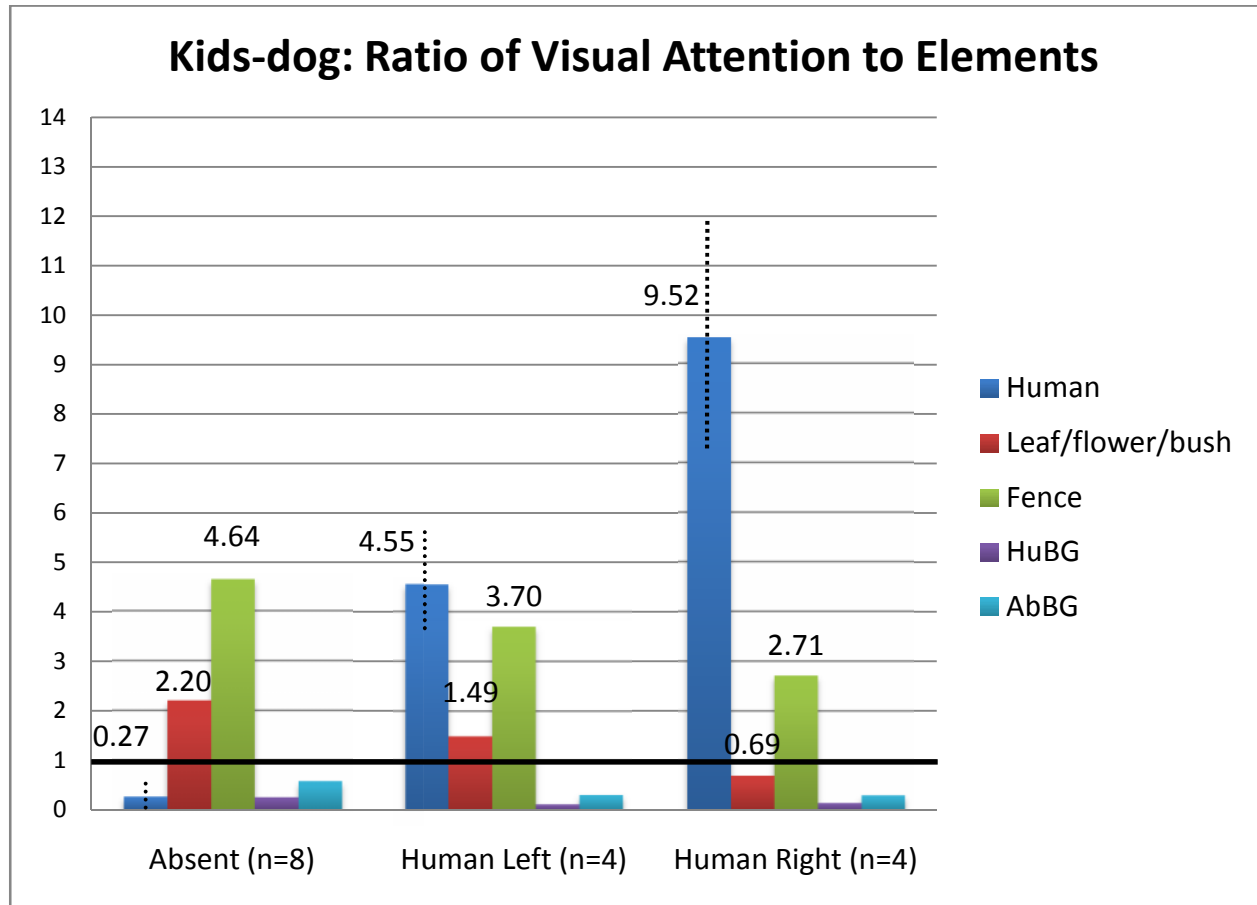


Figure G3. Kids-dog: Ratio of visual attention to elements.

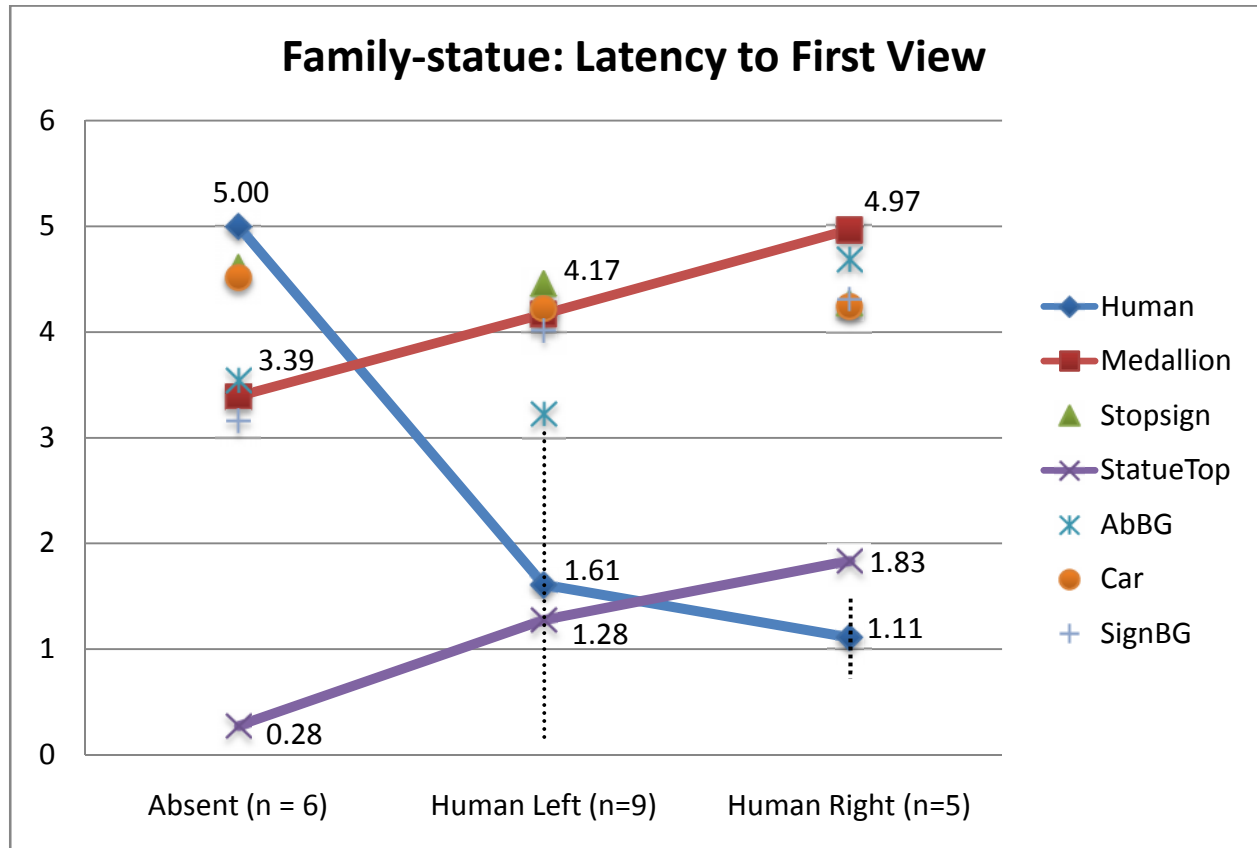


Figure H1. Family-statue: Latency to first view.

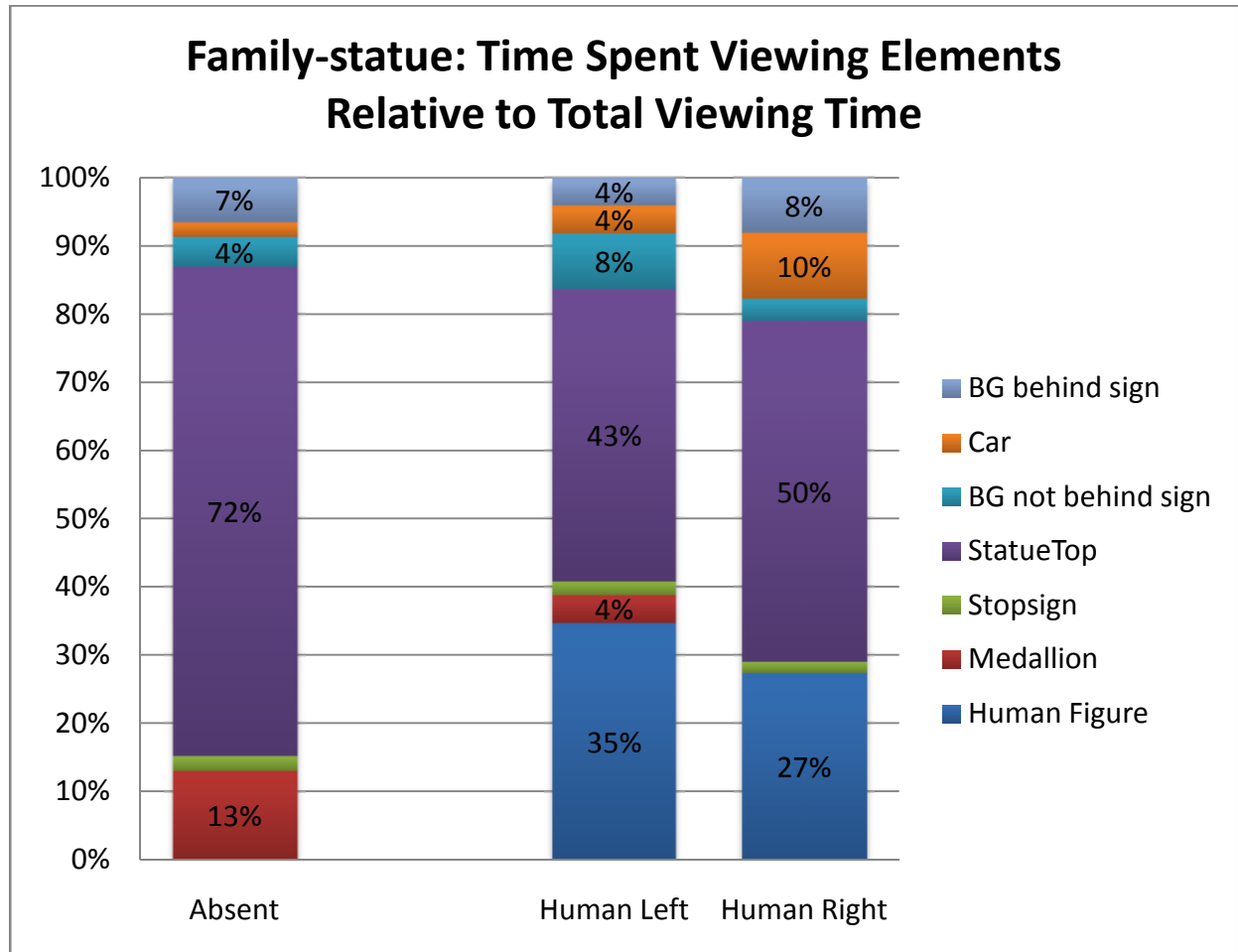


Figure H2. Family-statue: Time spent viewing elements relative to total viewing time.

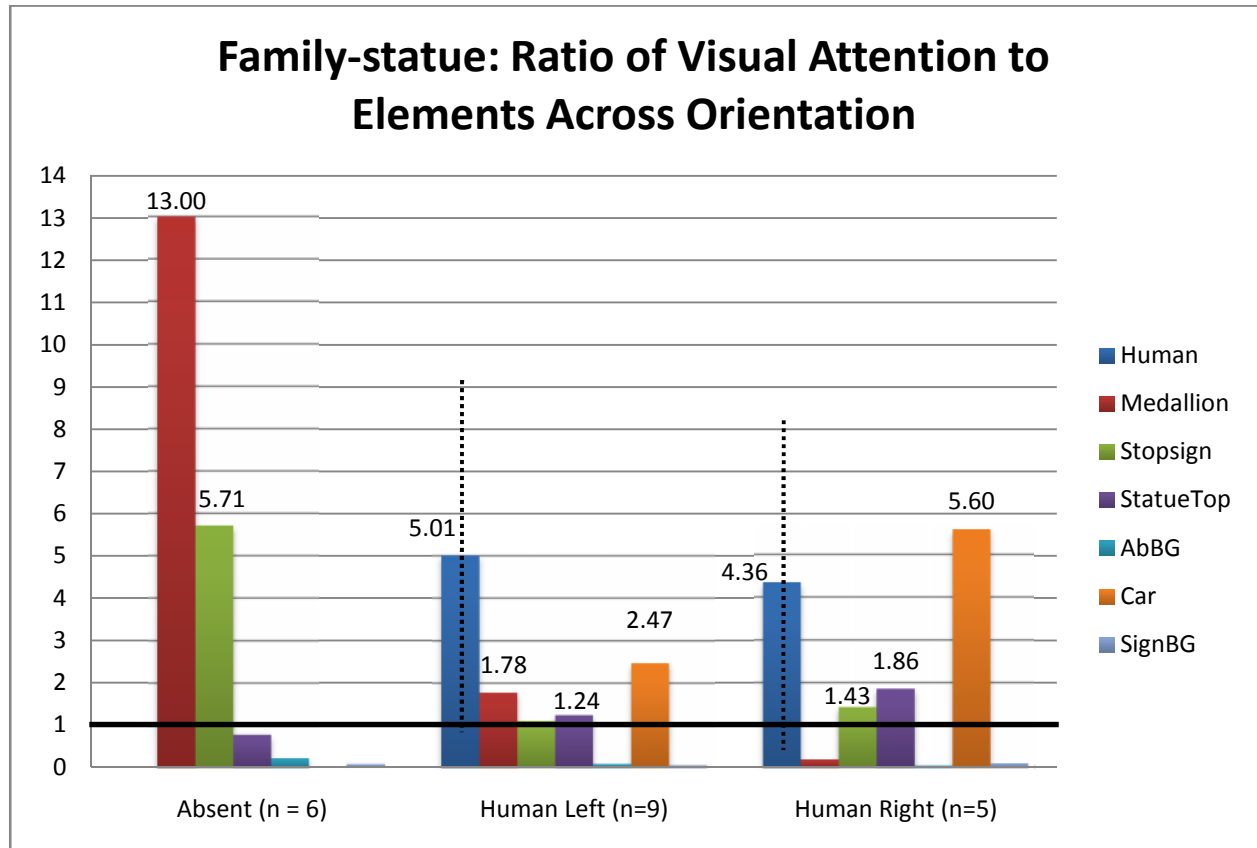


Figure H3. Family-statue: Ratio of visual attention to elements across orientation.

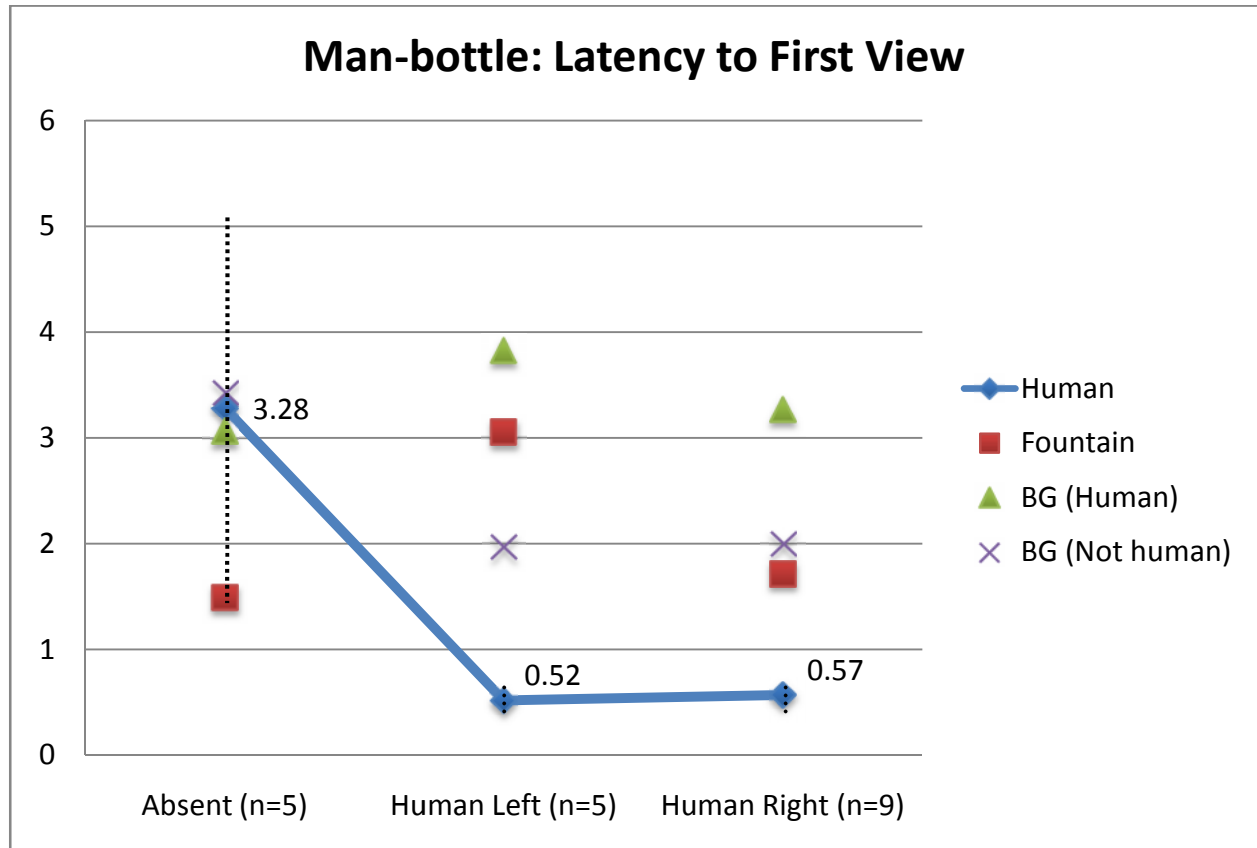


Figure I1. Man-bottle: Latency to first view.

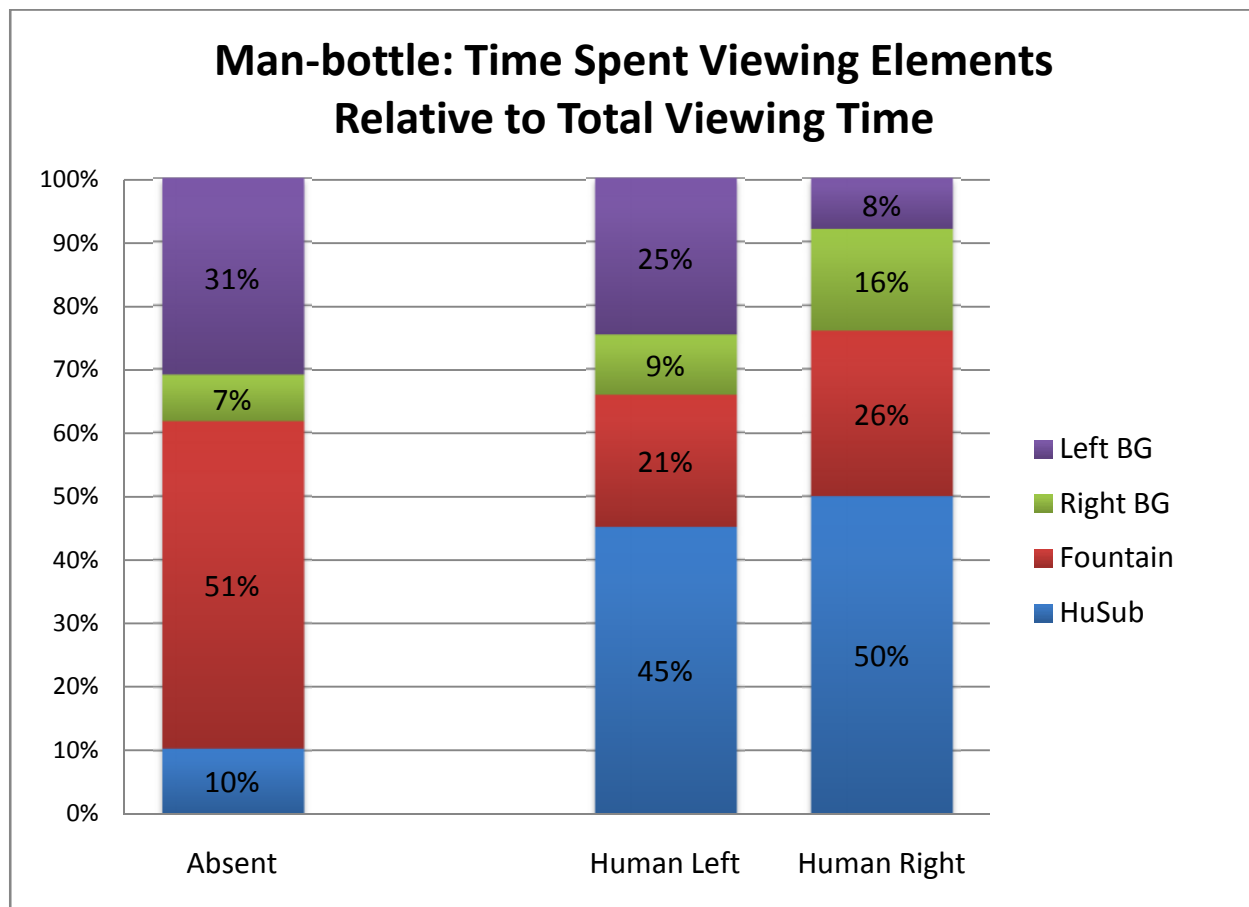


Figure I2. Man-bottle: Time spent viewing elements relative to total viewing time.

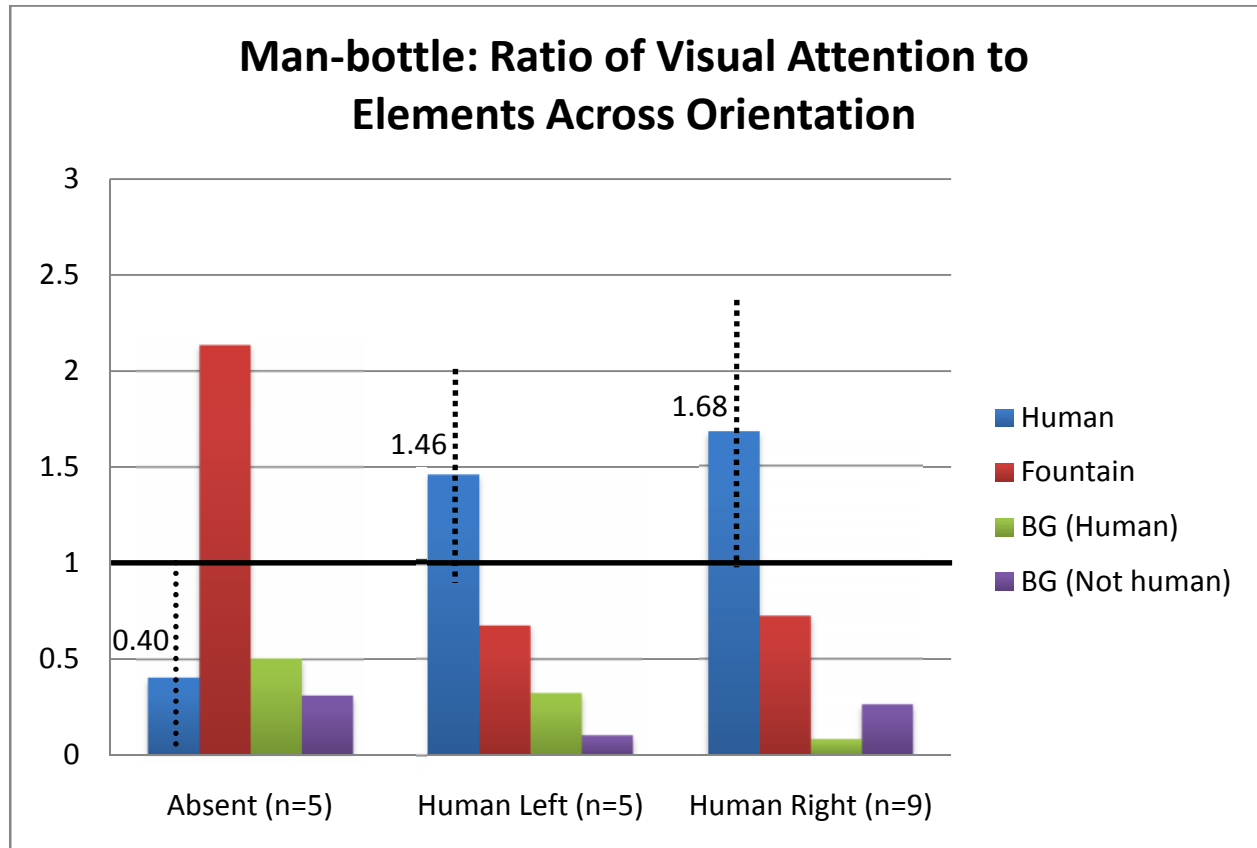


Figure I3. Man-bottle: Ratio of visual attention to elements across orientation.