How people look at pictures before, during, and after scene capture: Buswell revisited

Jason S. Babcock*, Marianne Lipps, and Jeff B. Pelz
Visual Perception Laboratory, Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, NY, USA 14623

ABSTRACT
The fact that some people take better photographs than others is reason enough to study how people perform this task. Insight into the picture-taking process can be accomplished by superimposing the subject’s fixation sequences over a live record of the scene. Monitoring eye movements provides a unique perspective because people are usually unaware of exactly what they looked at during the time of capture. While eye movements do not reveal the full cognitive processes underlying perception, they do provide an indication of where attention is deployed. Little is known about the relationship between how photographers compose an image and how they look at the picture once the image is available for display. Some of the earliest evidence of picture viewing comes from the work of Guy T. Buswell in his

Keywords: Eye tracking, visual perception, looking at pictures, image editing, fixation maps

1. INTRODUCTION

The fact that some people take better photographs than others do is reason enough to study how people perform this task. To gain further insight into the picture-taking process it is necessary to see what the photographer sees and to study what he/she pays attention to. An important question to ask is what exactly does the photographer look at while composing the image through the camera’s viewfinder? Further, what does the photographer look at in the real scene just before image capture, and how might these regions of interest compare to looking patterns made by the photographer when editing these images on a computer screen?

*Correspondence: jsb1623@cis.rit.edu
In 1935, Guy T. Buswell, a professor of educational psychology at the University of Chicago, published a book titled, *How People Look at Pictures: A Study of The Psychology of Perception in Art*. The work in this book is important to the history of eye tracking because it is the first thorough investigation to record and analyze the eye movements of 200 subjects while they viewed 55 photographs of objects ranging from paintings and statuary pieces, to tapestries, patterns, architecture, and interior design. Unlike the eye tracking technology used in early reading studies, the apparatus used in Buswell’s experiment was built specifically for the purpose of recording both the horizontal and vertical eye position during the course of image viewing. Buswell recognized that he was applying state of the art technology in order to gain new insight into what people did when they looked at pictures. In his introduction he states, “As is generally the case when a technique is first applied in a new field, this study possesses many of the characteristics of a survey experiment rather than one which tests carefully formulated hypotheses. The writer is in a much better position to set up such hypotheses now than at the beginning of the study (pg 17).”

In many ways the work published here parallels the work of Buswell in 1935. Specifically, a portable eye tracking system has been developed at the Visual Perception Laboratory at Rochester Institute of Technology that enables us to record the eye movements of subjects while they perform realistic tasks outside of the laboratory. Unlike the limitations imposed by eye tracking device used by Buswell, it is now possible to extend his question of how people look at pictures to how people take pictures. Further, with the advances in digital photography, it is now possible to study subject’s eye movements while they edit their pictures. With this in mind, the purpose of this paper is to present exploratory results from an experiment that recorded subject’s eye movements while they photographed and edited a series of digital images. However, in order to understand the motivation for this research, a discussion on visual acuity and eye movements is in order.

2. BACKGROUND

2.1 Eye Movements and Visual Perception

Evolution has equipped humans with a clever imaging system. Unlike a uniform CCD sensor in a digital camera, the eye’s retina is composed of two types of sensors; rods and cones. An important distinction between these sensors is their visual function. At intermediate levels of illumination, both rod and cone photoreceptors are active. As luminance levels decrease, rod sensitivity increases. Conversely, high luminance levels effectively saturate the rods so that only the cone photoreceptors are functioning. In the periphery of the retina, the rods greatly outnumber the cone photoreceptors. While a large distribution of rods is useful for seeing in low illumination conditions, visual acuity in the periphery is quite poor. At the center of the eye the cone photoreceptors are distributed in the region of the retina referred to as the fovea. Here several million cones are packed tightly together near the optical axis. From the center outward, the distribution of cones substantially decreases past one degree of visual angle (about the size of one’s thumbnail at arms length). Instead of pooling signals as in the periphery, each cone photoreceptor in the fovea reports information in a nearly direct path to the visual cortex. In the visual cortex, the fovea occupies approximately five times more neural tissue than the rods. For the most part, the spatial mapping of the retina (cone-center, rod-periphery) is analogous to the spatial mapping of the visual cortex. Given these characteristics, the fovea is responsible for high-resolution vision. Since the oculomotor system allows us to orient our eyes to areas of interest very rapidly, and with little effort, most of us are completely unaware that spatial acuity is not uniform across the visual field.

The temporal nature of eye movements is typically described as a combination of fixations and saccades. Fixations occur when the eye has paused on a particular spatial location in the scene. To reorient the fovea to other locations, the eyes make rapid angular rotations called saccades. On average, a person will execute over 150,000 gaze changes each day (Carpenter, 199?). This active combination of head and eye positioning provides us with a satisfactory illusion of high resolution vision, continuous in time and space. When performing everyday tasks, the point of gaze is often shifted toward task-relevant targets even when high spatial resolution from the fovea is not required. Since these ‘attentional’ eye movements are made without conscious intervention, monitoring them often provides us with a window into cognition (Findlay [19xx], Pelz et al. 1999). While eye movements do not expose the full cognitive processes underlying perception, most of the time they do provide an indication of where attention is deployed. The study of eye
movements is important because it is not exactly clear what mechanisms of vision, perception, and cognition help us plan the next eye movement.

2.2 Eye Movements and Picture Viewing

While Buswell was unable to quantify specific differences in eye movements between trained versus untrained artists, he did conclude that observers exhibited two forms of eye movement behavior. In some cases, viewing sequences were characterized by a general survey of the image, where a succession of brief pauses was distributed over the main features of the photograph. In other cases, observers made long fixations over smaller sub-regions of the image. While no two observers exhibited exactly the same viewing behavior; in general, people were inclined to make quick, global fixations early; transitioning to longer fixations (and smaller saccades) as viewing time increased.

When observers’ fixation patterns were plotted collectively for the same image, areas with a higher density of fixations often corresponded with semantic-rich regions. This result indicated that observers often fixated on the same spatial locations in an image, but not necessarily in the same temporal order. Further, these consistencies revealed that eye movements were not random, and were often influenced by the features of the image being viewed. Buswell also concluded that the “mental set” obtained from experimental instructions significantly influenced how people looked at pictures (pg 136).

A decade later, Brandt (1945) published another set of results based on the eye movement patterns of people looking at advertisements. His study also investigated the role of eye movements in learning strategies, as well as in the perception of art and aesthetics. Like Buswell, Brandt concluded that there were individual differences in eye movements, but in general, these behaviors were similar enough that certain “psychological laws” could be formulated (pg. 205).

In 1967, Yarbus, reported that eye movements were not merely visual reflexes tied to the physical features of an image. Instead he believed that the eyes were directed to areas in the image that were “useful or essential” to perception (pg 175). In his well known example, Yarbus recorded the scan-paths of subjects while they examined I.E. Repin’s, An Unexpected Visitor. During free-viewing, eye movement patterns across seven subjects revealed similar areas of interest. However, different instructions, such as estimating the material circumstances of the family, or remembering the clothes worn by the people, substantially changed eye movement patterns for one subject. In general, regions with the most information were likely to receive more fixations.

Mackworth and Morandi (1967) set out to investigate the relationship between image informativeness and fixation position (see Henderson and Hollingworth, 1998, for a review). Their research showed that fixation density was greater for semantically informative regions than non-informative regions. In their experiment, two color photographs were divided into 64 squares. One group of observers was asked to rate the informativeness of each square based how easily it could be recognized in a later session. A second group of observers (being eye tracked) were asked to examine both photographs with the intention of selecting the most pleasing one. The fixation frequency for a given square was found to correlate with rated informativeness, where highest fixation densities were spatially aligned with the most informative squares. Regions with low informativeness received few or no fixations at all.

In a global study of visual exploration and aesthetics, Molnar (1981) had two groups of students view eight classical pictures from Rembrandt to Chirico. In one group he told the subjects to view the pictures carefully as they would later be questioned about what they saw. These individuals were designated as the semantic group. He told the second group that of observers that they would be quizzed about the aesthetic qualities of the pictures (labeling them as the aesthetic group). Although the two groups were shown to have similar spatial fixation patterns, the duration of fixations was longer for the aesthetic group. This finding suggests that aesthetic motivation requires slower visual exploration than semantic visual exploration. [[[this seems dangerous: what does ‘aesthetic motivation’ mean? And can you explain what “slower visual exploration” means?]()] Post-experimental interviews revealed that the aesthetic group found the task difficult.

Nodine, Locher, and Krupinski (1993) later found that the composition of the image did influence how trained artists looked at paintings. In this experiment artists’ fixation durations were longer, and their eye movement patterns had a
tendency to focus on structural relationships between objects and backgrounds. For untrained viewers, fixation durations were shorter, and eye movement patterns focused mainly on pictorial elements that best conveyed objective reality.

The series of experiments started by Buswell in 1935 have focused on the perceptual and cognitive significance of eye movements relating to photographs, line drawings, and artwork already ‘captured’ by others. While these experiments demonstrated that observers tend to deploy their attention to similar regions in images created after image capture, they do not address the eye movements of observers before or during image capture. The focus of this research is to connect what we know about still picture viewing after image capture to oculomotor behaviors before and during image capture.

2. METHODS

2.1 Instrumentation

2.2 the tasks (capture, edit, survey)

2.3 the subjects]

The research described here relies on the ability to monitor a photographer’s eye movements during several phases of the photographic process, from scanning the scene before the photograph is captured through to editing the digital image on a computer. Special equipment is necessary to complete each component of the task. The first phase of the process is for the photographer to view an object or scene and decide what to photograph. We term that the exploratory phase. The second phase is when the photographer frames the photograph with the camera – the framing phase. The final phase is viewing the images. In this experiment the final phase required the photographer to crop the digital image on a computer screen – the editing phase. [[Jason – maybe the final experiment should also have a period when the person just looks at the images, without cropping. Maybe the task is just to rate the quality of each image?]]

To ensure that the experimental results were not unduly affected by the instrumentation necessary to track the photographers’ eye movements, special care must be taken to minimize constraints on the photographer during each phase of the photographic process, and to allow the photographer to behave as naturally as possible. Most traditional eyetracking instrumentation requires the observer’s head to be held stable by some form of head restraint. Even systems that are designed to allow free head movements require a tether between the observer and the eyetracker control instrumentation. In order to monitor the visual behavior of the photographer during the exploratory phase, we made use of a custom-made eyetracker developed at the Rochester Institute of Technology. To allow natural behavior during the editing phase, while collecting data about the photographer’s gaze position on the image during image editing, it was necessary to track the photographer’s head movements in addition to their eye movements. [[[Alternatively, describe the options: 1. stabilize the head 2. just collect video data, 3. track head and eye]]]

2.1 Instrumentation

2.1.1 Wearable Eye Tracking System

A wearable eye tracking system was used to record subject’s eye movements while they walked through the building taking photographs. The primary component of the tracker is of a pair of modified racquetball goggles as shown in Figure 1. The far left side of the goggles supports an optics module housing an infrared illuminator, miniature CMOS video camera (sensitive only to IR), and a beam splitter (used to align the camera so that it is coaxial with the illumination beam). An external first-surface mirror folds the optical path toward an infrared reflective mirror shown next to the nose bridge of the goggles. This mirror simultaneously directs IR illumination toward the pupil and reflects an image of the eye back to the video camera. When aligned properly, the illumination beam enters the pupil, retro-reflects off the retina, and back-illuminates the eye. The same phenomenon causes ‘red eye’ in photographs when a camera’s flash is aimed at the subject’s line of sight.
A second miniature CMOS camera is mounted just above the subject’s right eye to record the scene from the subject’s perspective. This provides a frame of reference to superimpose a pair of crosshairs that correspond to the subject’s point of gaze. Just above the scene camera is a small semiconductor laser and a two-dimensional diffraction grating. The laser/diffraction grating system projects a grid of points that are used to calibrate the subject’s eye movements relative to the video image of the scene. Since the laser is attached to the goggles, the calibration plane is fixed with respect to the head. The laser system greatly increases the accuracy of calibration and minimizes subject calibration time.

The backpack (Figure 2) carries a customized Applied Science Laboratory (ASL) 501 control unit. Eye and scene video out from the ASL control unit is piped through a picture-in-picture box so that the eye image gets superimposed onto the scene image. The combined video image is then recorded onto digital video camcorder as shown in Figure 3.
2.1.2 Integrated Eye and Head Tracking System

Horizontal and vertical eye position coordinates with respect to an LCD monitor were recorded using an ASL Model 501 eye tracker and Polhemus 3-Space magnetic transmitting system. Figure 4 shows the subject wearing the headgear shown in Figure 5. Because the system is based on NTSC video signals, gaze position on the LCD monitor is calculated at 60 Hz. Gaze position values can be averaged over a variable number of video fields to reduce signal noise. Eight video fields were averaged for this task, yielding an effective temporal resolution of 133 msec.
Gaze position (integrated eye-in-head and head-position and orientation) is calculated by the ASL using the eye-in-head signal described above and a head position/orientation signal from a Polhemus Fastrak magnetic field head-tracking system. The Fastrak system uses a fixed transmitter and a receiver attached to the eyetracker headband (Figure 4 and Figure 5). The ASL reports gaze position as the X-Y intersection of the line-of-sight with the LCD display. In order to calculate the gaze intersection point on the viewing plane, the position and orientation of the LCD display had to be measured with respect to a fixed point. The position and orientation of the display were defined by entering the three-dimensional coordinates of three points on the plane into the ASL control unit. The Fastrak transmitter is defined as the origin, and the distance to each of the three points (A, B, & C in Figure 6) is measured and entered manually. The gaze intersection point on the LCD display was collected to the laboratory computer for off-line analysis.

In addition to the data stream containing the gaze intersection on the LCD, the ASL creates a video record with a cursor overlay indicating gaze with respect to a scene camera attached to the eyetracker headband. Figure 7 shows a single frame from the video record.
Because the scene camera is not coaxial with the line of sight, calibration of the video signal is strictly correct for only a single distance. All gaze points are in the plane of the LCD display, and subjects typically do not change their distance from the display substantially, so the parallax error is not significant in this task, though it can be significant in tasks not constrained to a near-vertical plane. Note that the gaze intersection point calculated by the ASL by integrating the eye-in-head and head position/orientation signals is not affected by parallax -- the scene camera is used only during calibration when the distance to the scene is fixed. After initial calibration, the gaze intersection is calculated by projecting the eye-in-head position onto the display, whose position and orientation are known.

The ASL was calibrated for each subject before each trial session. Calibrating the ASL requires three steps -- 1) entering the position of the three reference points on the calibration plane as described above, 2) locating nine calibration points on the ASL controller, and 3) recording the subject's pupil and first Purkinje centroids as each point in the calibration target is fixated.

The ASL relies on a magnetic head-tracker to monitor the position and orientation of the head. The transmitter unit was mounted above and behind the subject's head. The transmitter contains three orthogonal coils that are energized in turn. The receiver unit contains three orthogonal Hall-effect sensors that detect the transmitters' signals. Position and orientation of the receiver are determined from the absolute and relative strengths of the transmitter/receiver pair. The position of the sensor is reported as the (x, y, z) position with respect to the transmitter, and orientation as azimuth, elevation, and roll angles.

2.1.3 Digital Camera

A Kodak DC-210 digital camera was used for image capture. The original size of the LCD panel was 1.63 x 1.20 inches, but for this experiment the LCD was masked off by a border of approximately 0.25 inches (see Figure 8). A mask closely resembling the appearance of the LCD was chosen so that subjects would be unaware of the alteration. The masking was done so that the photographers would capture more of the scene than was visible on the LCD. The extra image area would later give subjects a second chance to manipulate their composition on the computer display. Subjects were not notified about the image editing task until after they finished taking the nine photographs.
2.2 The Tasks

Subjects performed two primary tasks; capturing images with a digital camera, and editing their images on a computer. In addition, they completed a brief questionnaire at the end of the experiment.

2.2.1 Task 1: Digital Image Capture

After calibrating the wearable eye tracker, subjects were handed a 5.5 x 8.5 inch, double-sided, mock brochure describing the various types of research taking place in the Center for Imaging Science building. The inner spread of the brochure contained three empty rectangles meant to display, 1) a photograph of a person in the imaging science building, 2) a photograph of a large sculpture on the first floor, and 3) a photograph of the stairway in the building atrium. The scenes are referred to in the following text as “person,” “sculpture,” and “interior.” Figure 9 shows sample images of the three scenes.

The ten observers participating in this study were students and faculty from the Center for Imaging Science, so the observers were familiar with the people, offices and floors in the building. Participants were shown the digital camera and told that they would be using the camera to take pictures for a brochure about the Imaging Science building. Their instructions were to take three horizontal pictures of each of the three objects described above (totaling 9 images). Because it was not possible to track the eye movements directly through the camera’s viewfinder, all subjects were instructed to use the LCD panel to compose their images.

2.2.2 Task 2: Image Editing

After completing task 1 (on average, about 20 minutes) the subject reported back to the laboratory. The wearable eye tracker was taken off, and the person was informed that they would participate in a second task giving them the option to edit the images they had just photographed. Subjects were allowed to take a five minute break while the digital images from the camera were transferred to the computer and digital thumbnails were created.
A 20in x 30in Apple Cinema display, covering a visual angle of 30 degrees at 29 inches,

Subjects were calibrated and then told that they would view a set of thumbnails of the images they just took. Their task was to use the crop tool in Adobe Photoshop to improve the image. For this experiment, the crop tool was constrained so that the aspect ratio always matched the aspect ratio of the masked LCD viewfinder on the camera. Subjects were allowed to practice on a test image to familiarize themselves with the tool.

To start the actual task, subjects viewed digital thumbnails of the nine images they had just photographed. They were asked to select the best of the three images (one from each category) for final editing.

### 2.2.3 Questionnaire

Subjects filled out a questionnaire in order to gather information about the person’s skill level and experience with cameras and photography. Subjects were also asked to rate their satisfaction with the final, cropped image on a scale of 1 to 10. To determine whether they were aware that the LCD panel was cropped, the questionnaire included questions on whether they noticed anything awkward about the digital camera, or whether they noticed a difference between the images displayed on the screen and the photographs they remembered taking.

### 3. RESULTS

[[[outline: 3.1 – image capture results 3.2 Image editing results 3.3 combination of capture and editing ]]]

[[[In looking at the results, I think it is easier to follow (and more interesting to think of it this way) to look from the largest scale to the finest scale. That means first just compare the total time looking at the scene to the LCD time, averaged across the three scenes. That means looking at the sum of (object + surround) vs. camera. The second step is to split the data up into the three scenes (person, sculpture, and interior), again scene vs. camera. Finally, split it up further into object, surround, and camera.]]]

The results will be considered first for the capture phase, followed by the image editing phase, and finally the relationship between the two phases.

#### 3.1 Task 1: Digital Image Capture

One measure of behavior before and during image capture is to compare the total time that the photographer spends looking at the scene before taking a photograph to the time spent composing and taking the photograph. Figure 10 shows these values averaged across five subjects. The values for “scene” represent the sum of all gaze fixations on the object and surround before lifting the camera. The values for “camera” is the total time spent looking at the camera’s LCD panel. Note that the subjects were instructed to take three photographs of each scene (person, sculpture, and interior), so the times shown in Figure 10 indicate the time spent on all three photographs of each scene. It is evident from the Figure that subjects spent approximately the same amount of time looking at the scene before taking the photographs as they did framing and taking the photographs; about 42 seconds for all three pictures, or 14 seconds per photograph. [[be careful here, because there will be some confusion between the three “subjects” (or “scenes”) and the three photographs taken of each scene.]]

[[[It might be less confusing to present on the graphs the average time for each photograph, (1/3 the total used now) to give a better idea of what each action looked like. Maybe worth it for a full paper, but not here? We’ll have to figure out how to handle ]]].

[[[Q] Does the “camera” time also include the time looking at the image on the LCD after firing the shutter? That is probably to some degree due to the length that the...]]
image is shown by the camera. Q) Do people look at the image after capture until it goes off, or do they look away earlier?

Figure 10 Gaze duration before and during image capture for all photographs. Average gaze duration (in seconds) on the scene before lifting the camera, and on the camera as the photograph was framed on the camera’s LCD panel. Error bars represent one standard error of the mean for five subjects.

The analysis in Figure 10 averages values across the three subject classes; person, sculpture, and interior. Figure 11 addresses the question of how the subject being photographed influences photographers’ behavior. As is evident in the Figure, the subject has a significant influence on the time spent looking at the scene, but little effect on the time spent framing and taking the photograph. While it is difficult to draw inferences on image class based on the small sample, we note that the trend in the total gaze duration on the scene shows a marked increase with the extent of the subject. The subject in the person class [[do they spend most of the time looking at the person’s face?]] is smallest, and takes up a relatively small portion of the frame. The sculpture takes up a larger portions of the frame, and the interior still more.

Figure 11 Gaze duration before and during image capture for the three scenes. Average gaze duration (in seconds) on the scene before lifting the camera, and on the camera as they framed the photograph on the camera’s LCD panel. Error bars represent one standard error of the mean for five subjects.

A finer-grained analysis of photographers’ behavior was made by breaking down the total time spent looking at the scene into the time spent looking 1) at the object and 2) at the surround. The video record of the photographers’ eye
movements were analyzed to determine the total gaze duration on the primary object and on the surround in each scene class. This determination was straightforward in the person and sculpture classes, but more subjective in the interior class. The distinction is only clear after the photograph is composed and taken, as time spent on the object and surround regions reflects to some degree the decision-making process of what to include in the final photograph. Figure 12 shows the result of the analysis. The underlying cause of the increasing total time spent looking at the scene in Figure 11 is evident in Figure 12. Consider first the difference between the person and the sculpture classes. There is no significant difference between the time spent looking at the primary object in the two cases. The increase in gaze duration in the scene portion of image capture is due completely to an increase in the time spent looking at the surround. In the person class photographers spent approximately the same amount of time looking at the person as the surrounding regions (an average of ~10 sec for the three photographs of each class), while the gaze duration in the surround was nearly three times longer in the sculpture class (~30 sec). Unlike the case where the person evidently draws more attention, the photographer is more concerned with the ‘negative space’ when photographing the sculpture. This really means that we should try to make the same distinction during image editing – Marianne?][[we also need to acknowledge that we weren’t looking at typical case of amateurs photographing friends and family, but asking people to photograph a person related to the mock brochure.]] [for ‘future work’ - an open question is whether this difference is seen with professional photographers.]]

Next we consider the difference between the gaze duration on the scene in the sculpture and interior classes. The scene gaze duration is ~40% longer in the interior class. It is clear from Figure 12 that the entire increase is due to longer gaze duration on the object in the interior class. Recognizing the more subjective classification between object and surround gaze location in the interior class, it is important to consider whether the difference could be due to misclassifying object or surround fixations. This is not likely, as the average time the photographers spent viewing the scene regions (the sum of object and surround) increased 40% (~15 sec) between sculpture and interior classes. [[this isn’t clear … clarify or drop for now]]

Figure 12  Gaze duration before and during image capture for the three scenes. Average gaze duration (in seconds) on primary object, surround, and the camera LCD panel. Error bars represent one standard error of the mean for five subjects.

### 3.2 Task 2: Image Editing

#### 3.2.1 Duration and Number of Edit Windows

The videotaped records of each editing session were analyzed to determine the time spent editing each image, and the number of intermediate crop windows during the editing session. Unlike the dramatic differences between image classes observed during image capture, there was no significant difference in edit time or the number of edit windows. The trend toward longer edit times and more edit windows seen in Figure 13 is not significant. (P > 0.5).
While the average time spent per edit window was nearly constant across subject class (7.2, 7.1, and 8.0 seconds per window for person, sculpture, and interior, respectively), the variables were only weakly correlated ($R^2 = 0.22$), and differed between subject classes. Figure 14 shows the relationship for all three image classes; the slope of a linear fit pooled across classes was 3.3 seconds per window. Figure 15 illustrates how that value differed by class. The slope of the linear fits was 5.3 seconds per window for the person class, 4.4 seconds per window for the sculpture class, and only 1.0 second per window for the interior class.
The asymmetrical effect of scene class on capture and editing times is evident in Figure 16.

3.2.2 Fixation Density Analysis

The previous section focused on the time photographers spent editing their digital images, and on the number of edit windows before the final cropping was selected. In this section we consider the spatial distribution of gaze fixations during the editing process.
4. DISCUSSION

It is interesting to note the observed asymmetry between image capture and image editing. While there were very significant differences due to image class (person, sculpture, and interior) in the image capture phase, there was no significant difference in the editing phase. It may be the case the experimental design influenced the subjects to spend more time than they might otherwise have spent if they did not have the expectation that the final image would be used in a publication. [[[maybe future ask for family album?]]]

[[[future work]]]

The author wishes to extend the definition of composition given by Edwards further. By imposing a frame around some part of the three-dimensional world, the photographer or artist seeks to capture information about the scene. Regardless of whether the main object (or objects) belong there or not, to achieve “good” composition one must carefully fit together the positive shapes and the negative spaces so that the uninformative regions (negative spaces) do not interfere with the presentation of the informative (positive shapes)

4.1 Composition
In *Drawing on the Right Side of the Brain*, Edwards (1999, pg 120) describes composition as the way in which the artist chooses to arrange the components of an image into a particular frame. Specifically, composition deals directly with the placement of positive shapes (main objects of interest) and negative spaces (empty areas surrounding the objects of interest) into a set of bounding edges called the format. Typically, formats such as a sheet of paper, canvas, or photographic viewfinder, control how the artist chooses to fit the positive and negative spaces together into a unified whole. While experienced artists are keenly aware of the importance of format on composition, beginning artists are often oblivious to these boundaries. Edwards states that beginning art students tend to direct their attention almost exclusively to the objects or persons they are sketching. There is a tendency to regard the edges of the paper as nonexistent, almost as if the student perceived these objects in the real-world, where no explicit boundaries are defined. Not realizing the importance of format and its relationship to positive and negative spaces is often a key factor in poor composition.

In photography, the instant the eye looks through the camera’s viewfinder a format is imposed on the three dimensional world and the issue of composition is at stake.

This example is of particular interest because it supports Edwards’ observation that beginning art students tend to direct their attention to the objects they are drawing, paying less attention to the “wholeness” final composition.

4. ACKNOWLEDGMENTS

This work was supported in part by a grant from the New York State Science, Technology, and Academic Research program, and by Eastman Kodak.

5. REFERENCES

1. paper 1
2. paper 2
3. paper 3