# The effect of experimental instructions on the number of areas identified as important in photographic images

***Susan Farnand and Mark Fairchild; Rochester Institute of technology; Rochester, NY***

## Abstract

*A study is currently underway that is aimed at increasing understanding of the optimal design of pictorial stimuli in perceptual experiments. Evaluating the impact of image complexity on visual attention is of particular interest. Since this work centers on pictorial scenes, a variety of such scenes must be selected as stimuli. The experiments planned require scenes that are perceived to have at least five key areas of interest. Further, each of these must be able to be cropped to versions perceived to have three or four key areas of interest and one or two key areas of interest. The objective of the present experiment is to evaluate the impact of the experimental instructions on the number of key interest areas identified in each of the potential scenes and its cropped versions. The results of this experiment indicate that observer instructions have an impact on the number of areas observers determined to be important in the test images. These results have been used to select the scenes for subsequent work being conducted to evaluate the impact of scene complexity on how people look at images in perceptual experiments. In these experiments, fixation patterns will be evaluated with respect to the areas identified as important in the present study.*

## Introduction

Pictorial scenes produce a more complex visual experience than color patches. In an experimental setting, scenes present a greater opportunity than do uniform patches for observers' individual differences to significantly impact the process. Judd et al. (2011), in a study comparing fixation patterns of images at a wide range of resolutions, found that inter-observer fixation consistency depended on the complexity of the target image with simple images being more consistent. This also indicated that object recognition influenced fixation patterns; a finding supported by a study by Einhäuser et al. (2008) that found that objects contained within images better predict what catches observers' attention than early saliency. A study aimed at increasing understanding of the design of pictorial stimuli in perceptual experiments and the impact of the content of complex pictorial images on visual attention is currently in progress. As a first step in this study, the concept of what constitutes a complex image is being considered. An experiment was conducted to evaluate the number of areas perceived to be important in a variety of scenes, including images created by cropping the original scenes. In this experiment, an example image, showing a selection of image areas circled and numbered, was included in the observer instructions. The number of areas circled was varied for different observers to determine the impact of varying the experimental instructions on the number of areas identified as important by the observers.

The effect of task on eye movements when looking at complex images is well known. Torralba et al. (2006) found that task had a larger influence on fixation patterns than observer characteristics. Babcock et al. (2003), in their report on research regarding eye movements in varied psychometric scaling tasks, found strong evidence of task dependence. Ballard, Hayhoe, and colleagues have performed several investigations demonstrating that task overrides bottom-up saliency factors, see Rothkopf et al. (2007) for example. And Yarbus (1967) conducted experimentation in which he recorded the eye movements of an observer looking at a single image after receiving seven different instructions. While the eye movement data for this experiment show many similarities; the observer tends to look at faces in each case, for instance, there were substantial differences as well.

The study planned will involve analysis of eye movement patterns as a function of scene complexity. Including scenes exhibiting a range of perceived complexity, therefore, is important to the value of the study. Scenes consistently perceived to have at least five areas important to the information content of the image will be required. To effectively select scenes, more than just the researcher's opinion regarding image complexity will be needed. In this experiment, observers were asked to identify the important areas of 60 scenes by circling and numbering those areas. To illustrate the procedure, an example image was considered. However, this example may predispose the observers to identifying a particular number of important areas. To understand the effect of simply changing the example image in the observer instructions on how many areas the observers identify as important, this experiment was conducted with four different sets of experimental instructions.

## Experimental methodology

In this experiment, the 32 observers received one of four variations of the following written experimental instructions:

*Today you will be viewing 60 scenes. You have been provided paper copies of each scene. As each scene is presented, please circle and number the important areas of that scene on the corresponding copy. Important areas are those that convey information about the scene – it may be helpful to think about how you would describe the scene to someone else.*

*Start with the most important area (1), then the second (2), and so on until all important areas are identified. Some scenes might have few important areas and some might have many. An example is provided below. This is just one way this particular scene might be labeled. You might label this scene a different way. You might feel a different area is the most important, for example, or that <u>there are important areas in addition to those numbered</u>. You are to decide how many areas to number for each scene. There is no 'right answer'.*

*Thank you for participating in this study.*

 

***Figure 1.*** *The experimental setup. 'Paint Girl' image provided by Lexmark®.*


***Figure 2.*** *Examples used in the experimental instructions. Eight observers used the left image as a practice image, eight saw the center image, eight saw the right image, and the remaining eight observers saw both the center and right images.*

***Figure 3.*** *Examples of scenes used in the experiment. One example is included from the 'People' (top, 'Landscape' (second from the top, photograph by Dr. Mark Fairchild), 'Still Life' (second from bottom), and 'Composite' (bottom). The full scenes are on the left, followed by the mid-cropped images, and the two closely-cropped images. All scenes from the Corel® database at RIT unless otherwise noted.*

The observers were given a set of black and white printed copies of the images that they were shown and markers for labeling these prints. The images were projected onto a screen in a classroom in the Munsell Color Science Lab at the Rochester Institute of Technology, Figure 1. All observers were instructed to circle and number the 'important' areas in the images they were shown. However, eight observers were shown an example in which four areas were circled, Figure 2 – center. These observers received the instructions as written above. Eight observers were shown an example in which fourteen areas were circled, Figure 2 – right. A third group was shown both of these examples. For these two groups, the underlined clause in the instructions above was replaced with: *there are fewer important areas than those numbered.* (This clause was not actually underlined in the instructions provided and is only underlined here for illustrative purposes.) And the remaining eight observers were given a practice image with no areas circled, Figure 2 – left. For this group of observers, the words from "An example" through to the underlined words were eliminated. And the following sentence was added: *A practice image is provided below.* The author verbally described identifying four areas and fourteen areas as different ways the image might be labeled and then allowed time for the observers to label this image for practice, giving them an opportunity to ask questions and receive clarification. The order in which the different sets of instructions were used was randomized, though attempts were made to balance the gender and the facility with English of the observers in each group.

A total of 60 scenes were used in this experiment. The majority of the scenes were selected from the Corel® database available at RIT. An additional resource was publicly available pictorial ISO targets used in imaging standards since this is an important possible application for the results of the study. A digital capture of a painting from a fine art color reproduction study supported by The Andrew W. Mellon Foundation (Frey and Farnand, 2011) was included since experimental results from that study are available for reference. The remaining scenes needed to fill out the scene set were selected from the College of Imaging Arts and Sciences stock photo database and from the personal images of Dr. Mark Fairchild and the author.

Twelve scenes were selected for each of five typical categories of pictorial scenes used in perceptual experiments: (1) people, (2) natural landscapes that include foliage and sky, (3) still life scenes including fruit, vegetables, or flowers, (4) man-made objects such as yarn and buildings and (5) composite scenes containing natural settings or objects made of natural materials such as wood and metal along with man-made objects. An example of a composite scene is shown in Figure 3. Scenes in these categories are used in perceptual experiments so that scenes containing memory colors such as skin tones, blue sky, green foliage, and neutrals are well-represented. Typically, scenes containing wood or metal are often used as well. Scenes containing brightly-colored man-made objects are also included in perceptual testing so that a wide range of colors may be evaluated.

A variety of scenes are also being incorporated in this testing because subsequent testing will involve the impact of scene content on fixation. Several studies examining fixation locations suggest that scene category may have an important effect on attention. Testing by Jaimes et al (2001) and Babcock et al (2001) indicated that eye fixation patterns were relatively consistent within categories. And results of a study by Parkhurst et al (2002) suggested that different scene characteristics drew attention for different categories of images.

All potential images must be landscape-oriented since this will be required in future testing. This testing will also require that each scene have five or more areas of interest. And each of the selected scenes must be large enough so that they may be cropped to a scene containing one or two key areas while still being large enough (600x400 pixels) for viewing. Each of the 60 scenes were cropped such that the first cropping should result in approximately three or four key areas and the second cropping should result in scenes each having one or two key areas, see Figure 3, for example. This resulted in a total of 240 images. The images were divided into four sets such that each observer saw one rendition of each of the 60 scenes: 15 full images, 15 mid-cropped images, and 30 close-cropped images. Two observers given each set of the four sets of instructions evaluated each of the four image sets, resulting in a total of 32 observers.

The majority of the observers were recruited from a graduate Human Vision class. Extra-credit was offered for their time. Observers experienced in image evaluation (e.g. photographers or graphic designers) were avoided. Efforts were made to balance the observers' gender and facility with English for each set of images and each set of instructions.

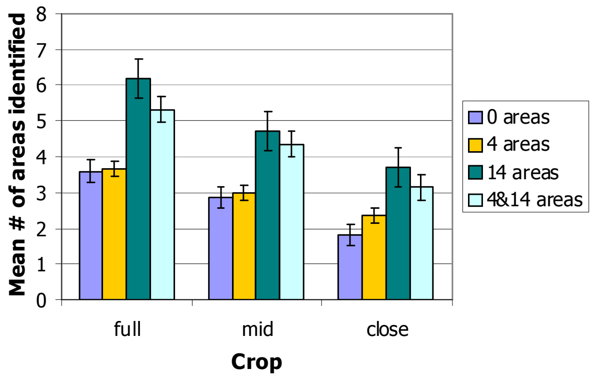


***Figure 4.*** *The mean number of areas identified by the groups of observers given each set of the experimental instructions for the full scene, mid-cropped, and closely cropped images.*

## Results and discussion

The mean number of areas identified for the full, mid-cropped and close-cropped scenes for each group of observers was calculated. The results are shown in Figure 4. These results show that the experimental instructions did have a statistically significant effect on the number of areas the observers identified as being important to the information content of the images. Observers who saw the example with four circled areas or received the practice image with no circled areas tended to identify fewer areas, on average, than the observers seeing either the example with fourteen circled areas or both the four and fourteen-area example images. This was true for the full scene, mid-cropped, and closely-cropped images. These two groups (0-area/4-area and 14-area/4&14-area) performed similarly to each other for the three different crops. Note that the group of observers receiving the instructions with no example received a verbal description specifically stating that circling four or fourteen areas were two possibilities. Yet this group performed significantly differently from the group receiving the instructions with the visual examples showing four and fourteen areas circled. The verbal description was insufficient to produce the same effect on the number of areas circled as the visual example.

Figure 5 shows the same data as Figure 4, but with the mean number of areas identified by the observers shown relative to the different instructions. These data indicate that there are clear differences between the numbers of areas identified for the full scene, mid-cropped, and closely-cropped images for the observers seeing either fourteen areas circled or both four and fourteen areas circled in the example images. In contrast, the observers seeing no areas circled or four areas circled did not identify a statistically significantly different number of areas for the full scene and mid-cropped images. Although they nearly all identified more areas in the full scenes, on average, they did not consistently identify enough more for the difference to be statistically significant.

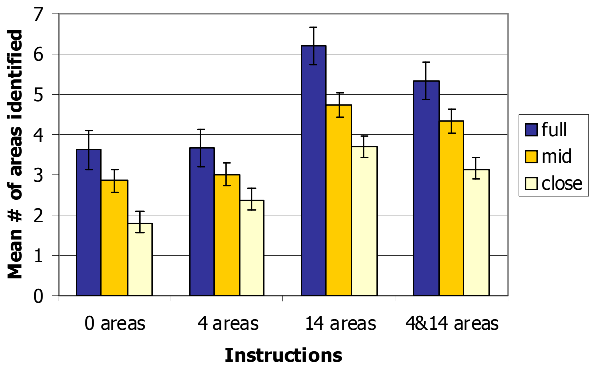


***Figure 5.*** *The mean number of areas identified by the groups of observers for the full scene, mid-cropped, and closely cropped images as a function of the number of areas circled in the experimental instructions.*

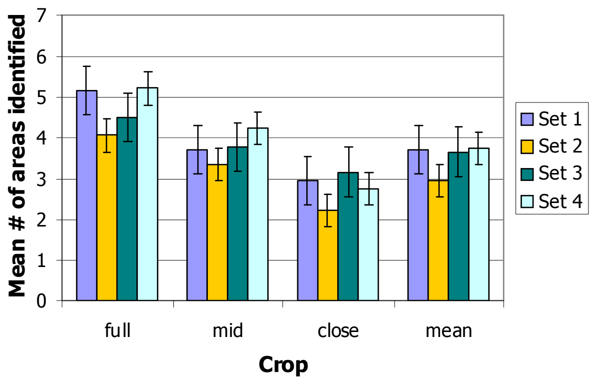


***Figure 6.*** *The mean number of areas identified by the observers seeing each set of images for the full scene, mid-cropped, and closely cropped images.*

In addition to receiving four different sets of instructions, the 32 observers also were shown four different sets of images, with eight observers seeing each set. (The eight observers seeing each set of images included two observers receiving each of the four different sets of instructions.) The mean number of important areas identified for each set of images is shown in Figure 6. These results indicate that the image set did not have a substantial impact on the number of areas circled. Only Set 2 exhibited any statistically significant differences. This set had lower numbers of areas identified, on average, than Sets 1 and 4 for the full scenes and Set 4 for the mid-cropped scene. These lower values were primarily the result of the observers receiving the instructions with the example having fourteen circled areas not showing the same increase over observers receiving instructions with the example with no or with four circled areas that was seen for the other three sets of images, Figure 7.

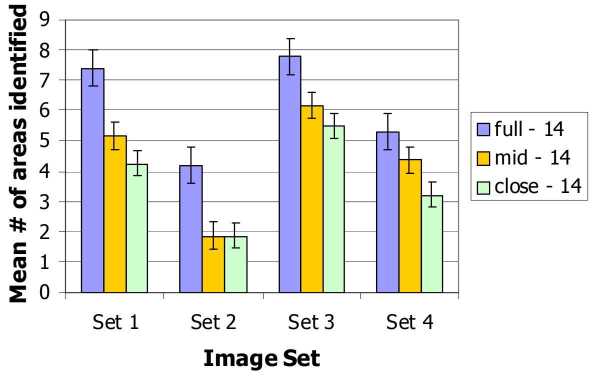


***Figure 7.*** *The mean number of areas identified by the observers who received the instructions with the example with fourteen areas circled for the full scene, mid-cropped, and closely cropped versions of each set of images. Note the lower values for Set 2.*

The intra-observer variability as determined by the standard deviation divided by the mean for each observer was evaluated, Table I. The observers receiving instructions with the example with no areas circled and with 4&14 circled areas have the same variation (.36 and .35, respectively). The observers seeing the 4-circled and 14-circled examples also have about the same variation, but at a lower level (.29/.31). This suggests that a more specific example produces more specific results. But, this lower level of variability is not necessarily desirable. The objective of the experiment was to determine the perceived number of important areas in the scenes provided. The instructions containing specific examples may be suggestive of the number of areas that observers should circle. This impact on variability may hint at undue influence of the specific examples in the instructions, rather than a consistent perception of complexity. The instructions showing no or two examples may provide a truer estimation of the perceived number of areas in each image.

**Table I: The average of standard deviation divided by the mean for the observers receiving each set of instructions**

| Instructions | Full scene | Mid-crop | Close-crop | Mean |
|---|---|---|---|---|
| 0 Circled areas | 0.42 | 0.28 | 0.38 | 0.36 |
| 4 Circled areas | 0.31 | 0.24 | 0.32 | 0.29 |
| 14 Circled areas | 0.34 | 0.25 | 0.34 | 0.31 |
| 4&14 areas | 0.37 | 0.32 | 0.36 | 0.35 |

The results of this experiment were used to select images for stimuli in an experiment examining the effect of image complexity on eye movements. In the course of this experiment, data will be collected regarding where observers look in the test images as well as the observers' perception of the complexity of the images. An approach similar to that of Einhäuser et al. (2008), in their study regarding image content and fixations, will be used. The approach involved asking observers to characterize the test scenes using up to five keywords. The data generated in this experiment will be evaluated with respect to the areas circled in this initial experiment to gain further understanding of perceived complexity.

## Conclusion

This experiment was conducted to determine the effect of varying the areas circled in an example image in the experimental instructions on the number of areas identified by the observers as important to the image content. The results indicate that the differences in the example image did have a significant impact on the number of areas identified. Observers who were shown an example having more areas circled tended to identify more areas as important. The observers seeing no areas or four areas tended to circle about 2-3 ½ areas, on average, while the observers seeing fourteen areas or both four and fourteen areas tended to circle about 3-6 areas, on average.

In additional testing, the instructions that included both the images with four and fourteen areas circled would be the recommended choice. These instructions yielded relatively consistent performance as well as distinct differences between the full scene, mid-, and closely-cropped images. An argument for the instructions with no example could also be made since this is the least suggestive approach. However, this set of instructions typically elicited more questions and consequently required a higher level of individual explanation leading to less consistency in the instructions ultimately being administered.

## References


[1] Babcock, J. S., Lipps, M. and Pelz, J. B. (2002). How people look at pictures before, during and after scene capture: Buswell revisited. In B.E. Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging V, SPIE Proceedings*, 4662, 34-47.

[2] Babcock, J. S., Pelz, J. B. and Fairchild, M.D., (2003b) Eye Tracking Observers During Rank Order, Paired Comparison, and Graphical Rating Tasks, PICS Conference Proceedings pp. 10-15.

[3] Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. Journal of Vision, 8(14):18, 1–26.

[4] Frey and Farnand (2011) Benchmarking Art Image Interchange Cycles: Final Report 2011, http://artimaging.rit.edu

[5] Judd, T., Durand, F. and A. Torralba. (2011) Fixations on low-resolution images. Journal of Vision, 11(4):1–20.

[6] Parkhurst, D., Law, K., and Niebur, E. (2002) Modeling the role of salience in the allocation of overt visual attention, Vision Research, 42, pp. 107-123.

[7] Rothkopf, C. A., Ballard, D. H., and Hayhoe, M. M., Task and context determine where you look, Journal of Vision 7(14):16, (2007) pp. 1–20, http://journalofvision.org/7/14/16/

[8] Torralba, A., Oliva, A., Castelhano, M.S., and Henderson, J. M. (2006) Contextual Guidance of Eye Movements and Attention in Real-World Scenes: The Role of Global Features in Object Search, *Psychological Review*, Vol. 113, No. 4, pp. 766–786.

[9] Yarbus, A., Eye Movements and Vision, Plenum Press, New York (1967)


## Author Biography


*Susan Farnand received her BS in physics from Cornell University and her Masters in Imaging Science from the Rochester Institute of Technology. She is currently working toward a PhD in Color Science at RIT. After beginning her career at Eastman Kodak, she moved to RIT where she works as a Research Scientist in the Center for Imaging Science. Her research interests include image quality, human vision and perception, and color science. She is a member of IS&T and serves as an Associate Editor for the Journal of Imaging Science and Technology, and has served as co-chair of the IQSP conference at EI and guest editor for the Journal of Electronic Imaging.*