

The Analysis and Prediction of Eye Gaze When Viewing Statistical Graphs

Andre Harrison¹(✉), Mark A. Livingston², Derek Brock²,
Jonathan Decker², Dennis Perzanowski², Christopher Van Dolson²,
Joseph Matthews², Alexander Lulushi², and Adrienne Raglin¹

¹ Army Research Laboratory, Adelphi, MD, USA

{andre.v.harrison2.civ,
adrienne.j.raglin.civ}@mail.mil

² Naval Research Laboratory, Washington, DC, USA

{mark.livingston,derek.brock,jonathan.decker,
dennis.perzanowski,christopher.vandolson,
joseph.matthews,alexander.lulushi}@nrl.navy.mil

Abstract. Statistical graphs are images that display quantitative information in a visual format that allows for the easy and consistent interpretation of the information. Often, statistical graphs are in the form of line graphs or bar graphs. In fields, such as cybersecurity, sets of statistical graphs are used to present complex information; however, the interpretation of these more complex graphs is often not obvious. Unless the viewer has been trained to understand each graph used, the interpretation of the data may be limited or incomplete [1]. In order to study the perception of statistical graphs, we tracked users' eyes while studying simple statistical graphs. Participants studied a graph, and later viewed a graph purporting to be a subset of the data. They were asked to look for a substantive change in the meaning of the second graph compared to the first.

To model where the participants would direct their attention, we ran several visual saliency models over the graphs [2–4]. Visual saliency models try to predict where people will look in an image; however, visual saliency models are typically designed and evaluated to predict where people look in natural images (images of natural or real world scenes), which have lots of potential information, subjective interpretations, and are not typically very quantitative. The ideal observer model [2], unlike most saliency models, tries to predict where people look based on the amount of information contained within each location in an image. The underlying theory of the ideal observer model is that when a person sees a new image, they want to understand that image as quickly as possible. To do this, the observer directs their attention first to the locations in the image that will provide the most information (i.e. give the best understanding of the information).

Within this paper, we have analyzed the eye gaze from a study on statistical graphs to evaluate the consistency between participants in the way they gazed at graphs and how well a saliency model can predict where those people are likely to look in the graph. During the study, as a form of mental diversion to the primary task, participants also looked at natural images, between each set of graphs. When the participants looked at the images, they did so without guidance, i.e. they weren't told to look at the images for any particular reason or

objective. This allowed the viewing pattern for graphs to be compared to eye gaze data for the natural images, while also showing the differences, in the processing of simple graphs versus complex natural images.

An interesting result shows that viewers processed the graphs differently than natural images. The center of the graph was not a strong predictor of attention. In natural images, a Gaussian kernel at the center of an image can achieve a receiver operating characteristic (ROC) score of over 80% due to inherent center bias in both the selection of natural images and the gaze patterns of participants [5]. This viewing pattern was present when participants looked at the natural images during the diversion task, but it was not present when they studied the graphs. Results from the study also found fairly consistent, but unusually low inter-subject consistency ROC scores. Inter-subject consistency is the ability to predict one participant's gaze locations using the gaze positions of the other ($n - 1$) participants [3]. The saliency model itself was an inconsistent predictor of participants' eye gaze by default. Like the participants, the saliency model identified titles and axis labels as salient. The saliency model also found the bars and lines on the graphs to be salient; however, the eye gaze of most participants rarely fell or focused on the line or bar graphs. This may be due to the simplicity of the graphs, implying that very little time or attention needed to be directed to the actual bar or line graph in order to remember it.

Keywords: Cognitive modeling · Perception · Emotion and interaction · Understanding human cognition and behavior in complex tasks and environments · Visual salience · Information theory · Statistical graphics

1 Introduction

Statistical graphs (SG) are visual representations used to represent high-order relations into different facets of data sets of a variety of types through the use of symbolic systems [6]. SGs present data as different types of images to take advantage of the advanced pattern detection capabilities of the human visual system and to allow an easy and consistent interpretation of information. SGs have numerous uses, including data exploration, structure investigation, outlier and anomaly detection, model verification, network flow, dimensional analysis, and more. Basic SGs include line graphs, bar graphs, and scatter plots. Line graphs and bar graphs are two of the most commonly used types of SG's and the type of SG used to present data can constrain or infer how that data is to be interpreted. In many fields, sets of SG's are used to detect and monitor complex types of information (widely known as "information dashboards"). Typically, the more specialized the use of the SG's is the harder it is for a novice to understand the information and the more time it takes to train a novice to be able to use these systems.

There has been substantial work in the design of SG according to human factors guidelines [7–10], which could help make these graphs more intuitive and reduce the learning curve of new trainees, but these guidelines are rarely used to create SG's in practice [7]. At the same time, some human factors guidelines can interfere with the intended application of the SG. It has been shown that features that make a graph more memorable can actually interfere with the comprehension of that graph [11, 12].

In this paper, we describe the results from a study aimed at studying the perception and comprehension of graph content, by seeing if participants could detect when the content in an SG changed. Participants studied a graph, and later viewed a graph purporting to be a subset of the data. They were asked to look for a substantive change in the meaning of the second graph compared to the first. In this paper, we analyzed the recorded eye gaze locations when the participants were examining the SG and the natural image in order to model the participants' perception of the SG by applying a visual saliency model to those same SG and comparing the results. Visual saliency refers to how likely items in an image are to attract a person's attention.

In Subsect. 1.1, we describe some of relevant work in the design of SGs, similar studies in the perception of SGs, as well as a brief background in modeling of visual salience. In Sect. 2, we briefly discuss the study that was conducted and the types of analyses done on the eye tracking data collected during the study. In Sect. 3, we discuss the results of the analysis on the patterns of eye gaze and the prediction of eye gaze using different saliency models. In the last section, we provide a short conclusion of the analysis and modeling done in this work.

1.1 Relevant Work

The Design of Graphs. There have been several papers on the design of SG [7–10]. The goal of these papers has been focused on how best to design graphs to convey an easily understood message to the reader, when a graph should be used to present information, and what type of graph should be used for different types of information [13, 14]. Research in the design of SGs has looked at how to make graphs more easily understood [8, 9], more memorable [11, 12], or how to control the aesthetic quality [15, 16].

There are many choices to be made when creating a graph, from the type of graph to use, the style of the text, the placement of text in the graph, as well as the types of graphical cues to use. Each of these choices can have an impact on the memorability of the graph and how easy it is to understand the primary and secondary meanings (if any) contained within the graph [7–10, 17–19]. In general, a graph with the simplest layout and the fewest number of variables and extraneous clutter is the easiest to understand, while a graph that is visually striking or has natural objects is more memorable.

Research into the memorability of graphs, by Borkin et al., has shown that the most memorable items in a graph are identified and encoded within the first 10 s of viewing [11, 12]. At the same time, they found that human recognizable objects in a graph improved a person's ability to recognize or recall the graphs. For graphs with less easily memorable features or objects standard text elements of a graph were used to recall and recognize those graphs, with title being the most popular point of recognition and recall [11, 12].

Understanding or interpreting the meaning of a graph starts with the type of graph used to represent the information. Certain types of graphs are selected based on the common interpretation of data when it is shown in a graph of that type [17]. Graph understanding is further colored by the salience of different graphical elements, the use of graphical cues, and the types of graphical cues used [17]. In general, graphical cues

have a strong impact on the interpretation and perception of an SG dominating attention over salient effects [17]. Work on the comprehension or understanding of graphs has shown that comprehending the trends or the effects that depend on interactions across graphs is difficult without training [18]. Untrained individuals don't see all of the information present in graphs, but what they do see they remember [18].

Work on the perception of SGs has focused on higher level descriptions of how the information in SGs could be encoded and represented within the brain [20, 21]. These descriptions are organizational; much like feature integration theory (FIT) by Treisman and Gelade [22], models of the perception of graphs like saliency models break up graph comprehension in terms of bottom-up and top-down processes. Most models, however, are not computational. Given a sample SG, these models can be used to tell a person how to evaluate how easily a graph can be understood or what level of knowledge a person needs to understand a particular graph [20].

Predicting Where People Look. Computational models of visual salience are a popular way to model visual perception. Visual salience is the aspect of visual perception that attracts a person's attention to a particular location in an image over others. Visual saliency models thereby predict where people look by identifying the most salient locations in an image and then assume that these locations are likely to be where a person will look. There are two main factors that influence the salience of an object or location in salience models, top-down factors and bottom-up factors. Top-down factors are learned or goal directed factors that affect visual attention. In short, they are external factors that influence where a person may look. In the realm of SGs, a common situation with a strong top-down factor is when a person is looking at an SG to answer a specific question. Top-down factors are also related to higher level processing tasks like object recognition, classification, and some aspects of segmentation [23–26]. Bottom-up influences are more sensory based; eye gaze is affected in this case by changes in contrast, texture, intensity, or color. Accounting for top-down influences in a model generally requires some sort of learning or training step that must employ a large database of positive and negative examples of objects or patterns. As a result, almost all current saliency models with top-down influences have only been trained on natural images. In the rest of this paper we focus only on bottom-up approaches to saliency. A benefit of bottom-up approaches is that while they have been developed and tested primarily on natural images their underlying assumptions and approaches are tied primarily to low level models of vision and so they can be generalized to most domains of vision.

Most bottom-up models of where people look take the natural approach to modeling attention by basing some or all of the components of their models on the neural circuitry involved in eye gaze [4, 27–31]. The theory of how attention works, on which these models are based, is from FIT by Treisman and Gelade [22] and Koch and Ullman [32]. Treisman and Gelade proposed that the extraction of features from a scene or image occurs in a parallel fashion extracting all the features from the image in one pass, while attention is a serial process and must jump from one region of the scene to another. Koch and Ullman expanded upon this theory by proposing the idea that the movement of attention from one part of a scene to another occurs in a winner-take-all (WTA) process and they first coined the term of “saliency map.” There were several

features initially thought to be extracted by Treisman and Gelade [22]; however, most biologically inspired bottom-up saliency models extract only three types of features intensity, color, and orientation at multiple spatial scales [4, 27, 28, 33].

Another approach that has been used to develop saliency models is to look at fixation from an information theoretic perspective. The underlying idea of these models is that when a person looks at a new image or scene they want to learn as much about the scene as quickly as possible. The locations that are gazed or attended to first are the ones that have the most information. In these models information is typically measured as entropy [2, 34–36] or self-information [5, 37, 38]. Other saliency models have been based in different areas of mathematics than information theory, like graph theory [3], and Bayesian estimation [39].

2 Materials and Methods

The objective of the study on SGs was to explore the manipulation of the information content in SGs. The approach taken in the study was to manipulate SGs (line graphs and bar graphs) in two ways. The first way the graphs were manipulated was by reducing the amount of information presented in each graph, thus graphs with 3–6 elements were modified into graphs with 1–3 elements. The other change was to sometimes change the message of the original graph, such that it no longer presented the same information. Each participant’s task was to detect if the information stated in the simplified graph was stated in the original graph (where “not stated” could imply a conflict or simply new information).

2.1 Procedure

The procedure followed within this study was to show each participant a very simple SG, then present the participant a series of distractors, afterwards the participant viewed a modified version of that same graph. The distractors were a series of natural images and a brief paragraph. For the images of SGs participants were given at least 30 s to study each image. For the natural images participants were told to free view the natural images for 3 s each. For the paragraph, participants were asked to read it, as they would have to answer questions about the paragraph, which they did after they were shown the simplified graph. For each trial, participants had their eye gaze tracked and recorded; however, while the eye gaze of the participants was tracked during the paragraph reading and answer sections they were not analyzed in this paper and so that aspect of the eye gaze data won’t be discussed further nor did we analyze the pattern of eye gaze when the participants viewed the modified graph. Each modified graph was only viewed by 4–5 of the participants.

2.2 Stimuli

The participants were shown 12 different SGs, 6 line graphs and 6 bar graphs. The graphs were chosen to be relatively simple and straight forward in meaning, such that if

the information was changed in the simplified graph the change would be clear. The natural images were selected primarily to serve as distractors to the main task, but they were selected from the MIT1003 dataset [40], which would allow the eye tracks to be compared to another natural image eye gaze dataset.

2.3 Participants

Twenty-four participants completed the study (20 men, 4 women) between the ages of 19–58 with a mean and median age of 38. Most participants were heavy computer users (at least 20 h per week). All but five participants said that they read or created graphs regularly (at least weekly).

2.4 Equipment

The participants' eye gaze was tracked using a Gazepoint GP3 eye tracker. The system, after calibration, tracked the participants' eye gaze for the rest of the study.

2.5 Analysis

Fraction of Valid Eye Tracks. The Gazepoint GP3 eye tracker is a desktop mounted eye tracker that tracks the left and right eye gaze of the person sitting in front of the device. It also maps the location each eye is looking into the pixel based coordinates of the display by using a 9-point calibration method where the viewer is prompted by the Gazepoint software to look at nine specific locations on the display. The nine different points of gaze allows the system to triangulate not just the user's eye gaze, but also the relative position of the user to the display and eye tracker in three dimensions. However, two main things can and did interfere with the eye tracker's ability to determine where on the screen the viewer was looking. (1) The viewer would sometimes look away from the screen, or at least outside the region where the SGs or images were presented. (2) The presence of eye glasses or any form of corrective lenses in front the participant's eyes. Glare from the computer screen or overhead lights could create specular reflections or highlights confusing the system's ability to locate the eye. Also the unknown prescription can refract the light in ways that can't be compensated for using the default calibration [41].

Observed Differences in Eye Gaze. It has become a typical finding that when analyzing the locations of gaze from a group of individuals who have been looking at a natural image without any stated objective there is a distinct central bias regardless of the content of the image [5, 29, 42]. This is present in every unguided eye tracking dataset using natural images. There have been several potential reasons given to explain this behavior. Commonly in eye tracking studies, participants are asked to focus on the center of the screen before the image is shown, therefore many eye fixations are found in the center. A similar justification is that when studying an image going back to the center is optimal from the perspective of minimizing the travel time

to any location in the image. Another explanation often given is that natural images used in studies are often taken by people and typically people take pictures so that the most salient object is located in the center of the image [43, 44]. We will compare the aggregated eye gaze locations for all users over all natural image against the aggregated eye gaze locations from the SGs to view any differences in viewing bias between the two classes of images.

Saliency Analysis. Saliency models are typically designed with the assumption that the image that will be analyzed is a natural one; with “natural” meaning an image that could be found in the real world. Thus, it could be argued that saliency models have no applicability to SGs. But the human visual system has evolved looking only at natural scenes, hence any image viewed by the human visual system may be initially processed as if it were a natural image. This is especially likely for bottom-up models of vision. So bottom-up saliency models may not be too far off in predictive accuracy when applied to SGs, in terms of identifying what might be salient. To evaluate how well saliency algorithms predict where people look in SGs when participants aren’t trying to answer a specific question, we analyzed the SGs and natural images in this study using several visual saliency algorithms.

The best predictor of where people will look in an image is where other people look when they see that same image, under the same circumstances. This is more formally known as the inter-subject consistency and if there are N participants in a study it is calculated by blurring the aggregated eye gaze locations of $N - 1$ of the participants into a 2D eye gaze map, treating it like a saliency map, and then using that map to predict the eye gaze locations of the N th participant. This is done for all N participants and then the average similarity value is calculated for all participants. Typically, the average inter-subject consistency represents the highest achievable accuracy possible by a saliency model. But the specific value has also been interpreted as a measure of the consistency of interpretation due to the consistency between the gaze locations [45]. We calculated the inter-subject consistency using two different saliency measures for the natural images and SGs used in this study. We also calculated the inter-subject consistency of the natural images using the eye fixation from the MIT1003 dataset [40], which is the original dataset the natural images came from.

There are several different similarity measures to compare a saliency map against a map of eye gaze [46]. In this paper, we selected two popular similarity measures, the receiver operator characteristic (ROC) and the Kullback-Leibler divergence (KL divergence). The receiver operating characteristic is a widely-used measure of the performance of a binary classifier. For succinctness, we have used the single-value version of the ROC and calculate the area under the ROC curve (AUC). The ROC curve compares the true positive rate vs. the false positive rate for each saliency map. By doing this it assesses the rank ordering of the values in the saliency map for each saliency model. The Kullback-Leibler divergence is a standard statistical measure for comparing the difference between two distributions; thereby comparing the overall pattern of fixation versus salience.

2.6 Results

Fraction of Valid Eye Tracks. A substantial portion of the eye gaze samples from the participant data was found to be invalid. This was partially due to the fact that three of the participants had no valid eye gaze information recorded at all. Discarding the users who had less than 10% valid eye gaze samples we found that 19.97% of the remaining eye gaze samples were invalid due to the eye tracker marking the recorded gaze as inaccurate, while, 37.52% of the gaze locations were not considered as the user's eye gaze didn't fall anywhere on the graph or image that was being shown. Overall only 13 participants had a sufficient number of eye gaze samples to be analyzed. From the remaining participants, the ratio of valid eye gaze samples had a mean and standard deviation of ($\mu = 0.4196$; $\sigma = 0.1563$). If the eye gaze samples are partitioned between whether the participant was looking at an SG vs. a natural image we find that 60.57% of the eye gaze samples are valid for the natural images and 38.86% of the eye tracks are valid for the SGs. If we further restrict the eye tracks of the SGs to the first 3 s 21.53% of the eye tracks are valid. Overall this means that there are 59,679, 11,656, and 188,986, valid samples of eye gaze when participants were looking at the natural images, at the SGs for three seconds, and at the SGs for the full study period, respectively.

It is of note that 13 of the 24 participants reported wearing some sort of prescription corrective lenses and 3 additional participants reported that they use non-prescription reading glasses. However, whether the participants were wearing some form of corrective lenses during the study was not recorded and not every participant who was eliminated from the analysis wore some form of corrective lenses. To determine the correlation between invalid eye gaze samples and corrective lenses we computed three binary dummy variables for eyewear:

1. Any form of eye correction = 0 or "Poor" and No correction = 1 or "Good"; this variable is called "No Correction"
2. Prescription eyeglasses = 0 or "Poor" and anything else = 1 or "Good"; this variable is called "No Script Glasses"
3. Prescription or Non-prescription eyeglasses = 0 or "Poor" and anything else = 1 or "Good"; this variable is called "No Glasses".

Any form of correction includes contact lenses, prescription eyeglasses, and non-prescription reading glasses. We suspected based on the literature and conversations with experts that prescription eyeglasses would create challenges for the eye tracker, because the frames and (often) thick lenses create unexpected reflections in the images that the eye tracker uses to locate the pupils. By similar logic, we suspected reading glasses to create challenges as well, since they have frames (and despite the lenses generally being thinner than prescription glasses). We did not expect contact lenses to create challenges for the eye tracker, since the lens is so close to the cornea. Our three dummy variables include (respectively) all three corrections (hypothesis: too restrictive), just prescription eyeglasses (hypothesis: too permissive), prescription or non-prescription glasses (hypothesis: best predictor). All these hypotheses are supported by the following analysis.

We computed analysis of variance (ANOVA) for the dummy variables. This also showed that the three dummy variables encapsulating the use of eyewear were a significant factor in the percentage of valid eye gaze data (Table 1). Note that the generalized effect size (GES) for each variable is considered large.

Table 1. Analysis of variance (ANOVA) for dummy variables regarding eyewear and the percentage of valid eye gaze data. Each (binary) dummy variable shows a significant main effect on the percentage of valid data, indicating that, in particular, eyeglasses – whether prescription or not – seemed to create challenges for the eye tracker to acquire valid eye gaze data. GES is generalized effect size; it is considered large for each variable, implying that each variable could explain a significant amount of the variance in percentage of valid data. It appears that the best hypothesis is that any form of eyeglasses created challenges for the eye tracker, whereas including contact lenses wear as a predictor of poor performance explained less of the variance.

	F (1,22)	P	GES	Mean ratios	
				Good (SEs)	Poor (SEs)
No correction	4.4040	0.0481	0.1734	0.38 (0.08)	0.19 (0.05)
No script glasses	9.3274	0.0060	0.3076	0.36 (0.06)	0.11 (0.05)
No glasses	9.6320	0.0054	0.3144	0.40 (0.07)	0.14 (0.05)

Observed Differences in Eye Gaze. Aggregating the eye gaze locations from all of the valid participants across all of the images shown creates a picture of the viewing patterns and biases people employ when viewing imagery. In Fig. 1a, a blurred version of the aggregated eye gaze locations from all of the valid participants over all of the natural images used in this study shows the aforementioned central bias typically found in unguided eye gaze studies of natural images. Repeating that procedure for the SGs creates a very different pattern of eye gaze, Fig. 1b. The pattern of eye gaze on SGs shows that in this study there was consistently more than one location that drew a participant’s attention. The fixation pattern for the SGs shows clusters of fixations near the edges of the image, falling where text for the graph is usually found: the title, the y-axis label, the x-axis label, and the legend. Strangely however, the actual content (the line in the line graph or bars in the bar graph) does not seem to attract as much attention as does the text on the x and y-axis, the title, or the legend. Part of this is likely due to the simplicity of the graphs. There are few data points in the line or bar graphs and there are no secondary factors or interactions to interpret. If the participants are viewing the graphs for upwards of 30 s, simple bars and lines are unlikely to require much attention. There is slightly more information in the text and the numbers for each axis, especially since the task was to study the graph in order to identify differences later. Patterns of eye gaze are typically different in free viewing tasks compared to memorization tasks. In memorization tasks there is a stronger top-down influence, there tend to be more eye fixation locations, and the average distance between fixation locations is larger [47]. The increased length of time the participants viewed the SGs meant that both bottom-up and top-down factors were guiding the participants viewing behavior. In the first 3 s of viewing the SG bottom-up effects are likely to dominate the eye gaze pattern. Samples from the first 3 s of viewing the SGs seem to primarily be drawn to

the x-axis label for both bar and line graphs, Fig. 1c. Two of the implications of this behavior are that eye gaze is largely controlled by information content and initial fixation or optimal search strategies are not the main cause of central bias in natural images [43, 44].

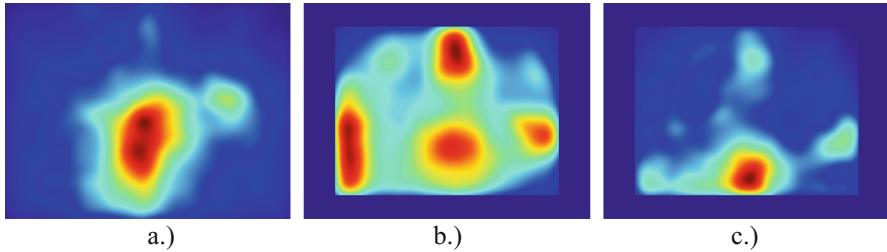


Fig. 1. Aggregated eye fixation for all of the participants. (a) Shows the aggregated eye fixations for all of the participants when they were looking at a natural image. (b) Shows the aggregated eye fixations of all of the participants over all the SGs for the entire time they were looking at each graph. (c) Shows the aggregated eye fixations of all of the participants over all of the SGs for the first three seconds that they were looking at the graph. The SGs were smaller than the natural images and so there is an empty border around the aggregated eye fixations for the SGs.

Saliency Analysis. The calculated inter-subject consistency for the SG and natural images used in the study on SG has fairly low AUC values, Table 2. As a point of reference the calculated inter-subject consistency of the participants' eye gaze patterns when looking at natural images from the SG study compared to the MIT1003 study shows a substantial drop in consistency using the AUC measure. The differences in the inter-subject consistency of the natural images shows that, assuming the different pools of participants are interchangeable, the substantially lower scores from this study are likely a result of the dropped eye gaze samples. The substantially lower inter-subject consistency scores from this study are likely due to participants with a smaller percentage of valid eye gaze samples or participants whose valid eye gaze locations are close together. If locations where most of the other participants directed their gaze didn't coincide with the eye gaze location of the remaining participant due to dropped samples, then that location is marked as a false positive lowering the overall consistency score and changes the rank ordering of the salient and non-salient locations.

Table 2. Comparison of the inter-subject consistency of the eye gaze patterns analyzed by this paper broken out by image type (natural image and SGs) and the per image trial length (the entire trial vs 3 s). Also on the table is the inter-subject consistency using the eye gaze patterns from the MIT1003 dataset.

	AUC	KL divergence
Statistical graph (all)	0.6460 ($\sigma = 0.0789$)	1.2864 ($\sigma = 0.6524$)
Statistical graph (3 s)	0.6216 ($\sigma = 0.0397$)*	5.0776 ($\sigma = 1.3348$)*
Natural images	0.6606 ($\sigma = 0.0888$)	3.2033 ($\sigma = 1.7713$)
Natural images (MIT)	0.8674 ($\sigma = 0.0363$)	2.7507 ($\sigma = 0.7234$)

Conversely, the inter-subject consistency values for the KL divergence seems to show no real difference when comparing the eye gaze patterns over the natural images. This suggests that the KL divergence values for the SGs represent a true upper bound of the best a saliency model could be expected to do in terms of predicting where a person would look when memorizing an SG. This also suggests, that the KL divergence computation is more invariant to the holes in a participant’s eye gaze samples. This seems like a reasonable conclusion as the KL divergence basically compares the difference in distribution of the saliency values where a participant looked against the saliency values where they didn’t look. So even if there is missing data so long as the separation in saliency values between gaze locations and non-gaze locations is wide enough the KL divergence values won’t be affected. This also explains why the KL divergence value is so much lower when the KL divergence is calculated using all of the eye gaze samples from the SG trial vs. the first 3 s of the trial. There are more eye gaze locations as the number of eye gaze samples is increased, increasing the variability in the saliency values at the gaze locations thereby reducing the differences in the distributions.

We applied several different saliency models to the images of the SGs (the Itti model) [4], the graph based visual saliency model (GBVS) [3], and the ideal observer model (ioM) [2]. In the evaluation of the saliency models we also include a saliency map that is simply a Gaussian kernel centered in the middle of the image. The accuracy

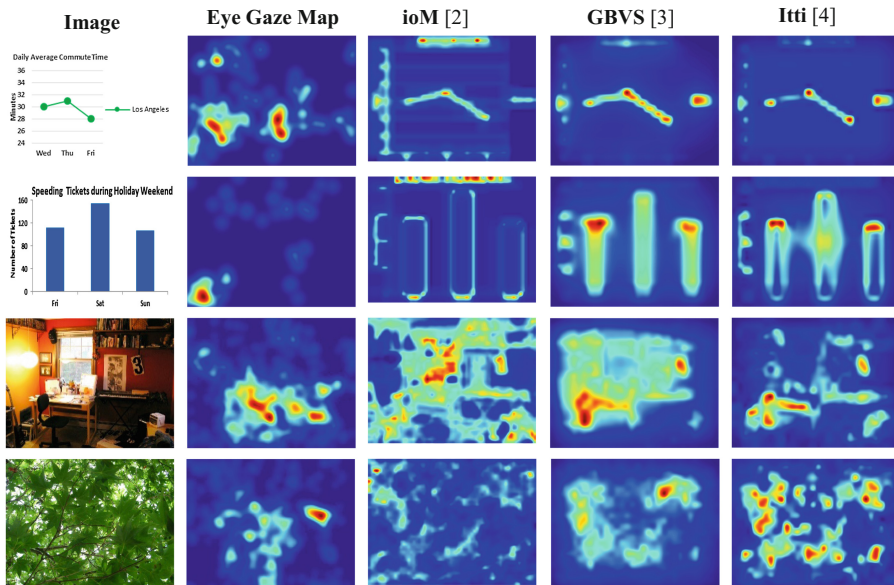


Fig. 2. Qualitative comparison of the saliency maps from the different saliency models. The 1st column, the leftmost column, shows the original images used within the study. The 2nd column shows the aggregated eye fixations for each image. The 3rd, 4th and 5th columns shows the saliency maps generated by the ideal observer model (ioM) [2], the graph based visual saliency (GBVS) model [3], and the Itti et al. model [4].

of the Gaussian kernel sets a minimum level of required predictive accuracy since the predictions are solely based on central bias and don't change with image content. Each saliency model generated a saliency map for each image, showing a 2D probability of the likelihood that each location would attract someone's attention, Fig. 2. These maps were compared against the aggregated maps of eye gaze for each SG, Fig. 1. Consistently across all of the saliency models the saliency maps of SGs find not just the text areas salient, but they also identify the graph and bars as salient. However, it is unclear what impact this has on each measure of accuracy.

In the first evaluation, we compared the ability of the saliency models to predict eye gaze when people looked at natural images, at the SGs for 3 s, or the SG for the entire memorization time (>30 s) Fig. 3. We specifically compared the output of the saliency models against the gaze patterns from the first 3 s of the SG trial for two reasons: (1) The participants could only look at the natural images for three seconds, so using only the first three seconds of the eye gaze patterns for the SGs makes for a more appropriate comparison. (2) In the first three seconds, bottom-up influences are still a strong predictor of eye gaze.

Across all of the models the predictive strength of the saliency models for the AUC measure follows an expected progression, Fig. 4. Every model has a higher accuracy when applied to a natural image, a slightly lower accuracy predicting eye gaze in the first 3 s of viewing an SG and even less accuracy predicting eye gaze for the entire time participants were viewing the SG. The results of the KL divergence however, shows a more unusual pattern. Each saliency model still has the highest accuracy predicting eye

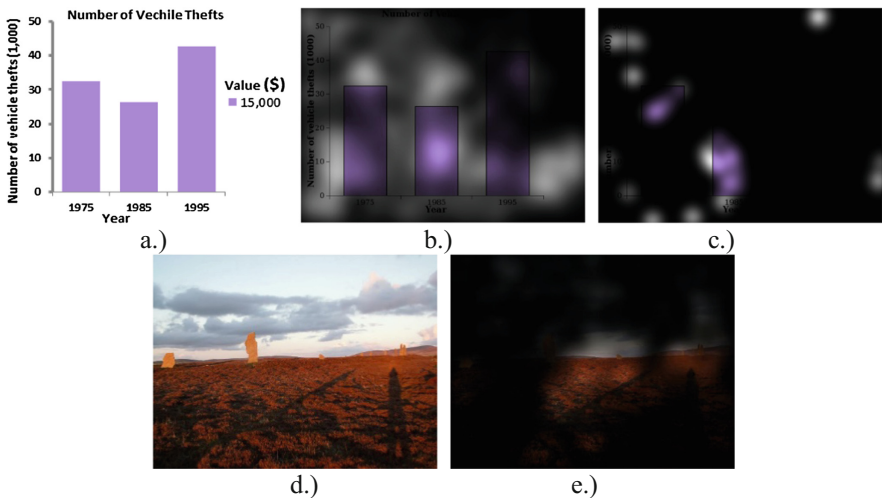
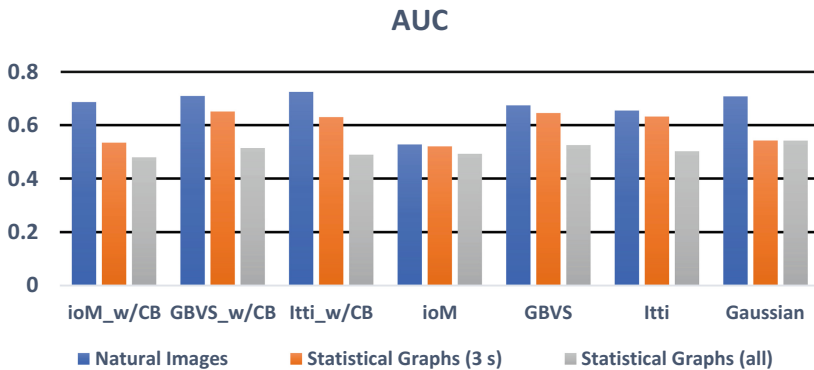


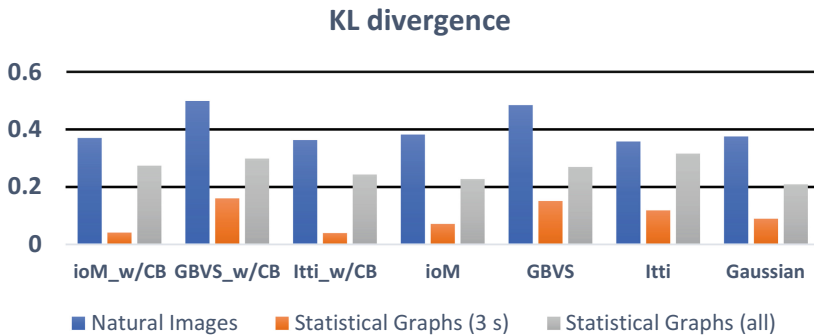
Fig. 3. (a) Sample SG used in the study. (b) Aggregated eye fixations of all the participants overlaid over the original graph. (c) Shows the aggregated eye fixations of all the participants during the first three seconds of viewing the graph. (d) Sample natural image used in the study. (e) Aggregated eye fixations of all the participants overlaid over the original image. Each eye fixation location is blurred to $\sim 1^\circ$ of visual angle. This visualizes the approximate accuracy of the GazePoint eye tracker and the coverage area of the fovea of the average participant.

gaze on for natural images, but the accuracy in predicting eye gaze from the entire SG trial is higher than for the first three seconds. A possible reason is that with the longer viewing time the pattern of eye gaze is more spread out over the entire image. When the KL divergence calculates the distribution differences between saliency map values at eye gaze locations vs non-eye gaze locations the longer eye gaze time reduces the number of non-eye gaze locations that fall on a high saliency map value.

Finally, a more anomalous pattern shown in Fig. 4a and b is that that even though the saliency models have a higher predictive accuracy when applied to natural images they have the same or a lower similarity score than the Gaussian kernel using the AUC similarity measure. However, all of the saliency models have a higher similarity measure than the Gaussian kernel when predicting eye gaze locations during the first



a.)



b.)

Fig. 4. Comparison of the predictive accuracy of the ioM, GBVS, and Itti saliency models when applied to natural images, and SGs using different similarity measures. (a) Uses the area under the curve of the receiver operating characteristic similarity measure. (b) Uses the Kullback-Leibler divergence similarity measure. Results are shown for the saliency models with center bias (left three groups of bars) and without center bias (right three groups of bars). The results of the Gaussian kernel are typically considered minimum necessary performance for minimum accuracy.

3 s of the SG trial, but only for the KL divergence similarity measure. When predicting the eye gaze for the entire SG trial most of the saliency models have a higher AUC score compared to the Gaussian Kernel, but in this case, all of the tested saliency models had a higher KL divergence similarity measure than the Gaussian kernel.

We also compared how the predictive accuracy of the saliency models used in this paper changed when eye gaze locations from the MIT1003 dataset were used, Fig. 5. For both similarity measures, an improvement in the predictive accuracy occurred in all of the saliency models we tested. Another improvement that occurred is that the Gaussian kernel no longer matches or outperforms the predictive accuracy of the saliency models using the eye gaze information from the MIT1003 dataset. Hopefully, this means that by using a more precise eye tracker, the predictive accuracy of the tested saliency models will also improve, using the AUC similarity measure when applied to SGs. What all of this most likely means is that bottom-up saliency models can with some accuracy, predict where people first look at simple SGs, but a more precise eye tracking system is necessary to definitively show this.

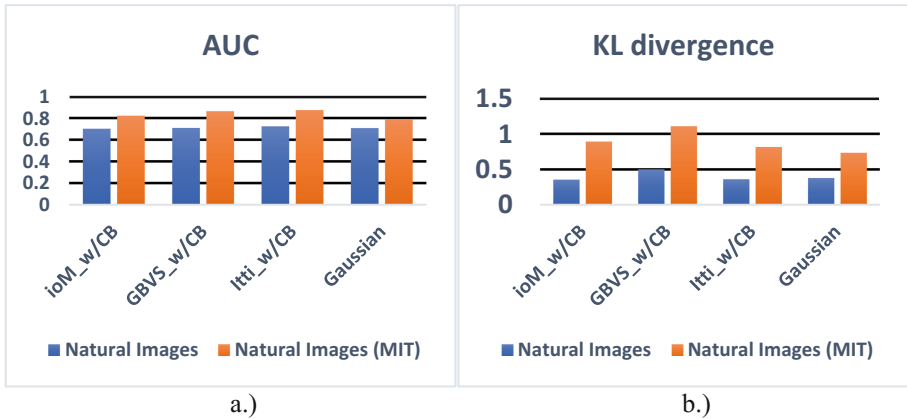


Fig. 5. Comparison of the predictive accuracy of the ioM, GBVS, and Itti saliency models with center bias when applied to the natural images in this study, but evaluated using the eye gaze locations from this study vs. the eye gaze locations from the MIT1003 dataset using different similarity measures. (a) This graph uses the area under the curve of the receiver operating characteristic similarity measure. (b) This graph uses the Kullback-Leibler divergence similarity measure.

3 Conclusion

Within this paper, we have analyzed the eye gaze of participants while they memorized SGs and natural images. We have shown that a person's viewing pattern when looking at simple SGs is different from their viewing pattern when looking at natural images. When a person looks at a natural image there is a strong unimodal bias near the center of the image. This strong central bias allows a person's gaze to be predicted, at higher

than chance levels, using just a Gaussian mask centered in the middle of the image. The viewing pattern for the SGs analyzed is far more multimodal. For the graphs used in the study that we analyzed there were distinct attractors of the participants' gaze that was consistent across graph type and graph content. These attractors of attention are focused on the graph axes, title, and to a lesser degree the graph legend. These results are consistent with the results of other studies that have looked at the eye gaze patterns of individuals when they looked at natural images or SGs [11, 48].

Many visual saliency models try to predict where people will look in an image by trying to model some component of the human visual system. These models are typically only applied to natural images, but using bottom up principles of vision SGs can be processed like any other image. Based on the performance of the different saliency models in the results section it does seem like saliency models can predict where people will look in a graph, for the first few seconds at least, with accuracies better than chance and better than just assuming the center of the image is important. One hindrance to improving the predictive accuracy of saliency models is likely due to the fact that SGs are generally more structured than natural images. There are certain locations, regardless of the type of graph, that are likely to contain relevant information and thereby be salient. This bias is similar to the central bias present in natural images and possibly this bias can be incorporated into a saliency model for SGs. However, the graphs that were analyzed in this paper had titles and axes roughly in the same location. For this bias to be used on more generic SGs it would be necessary to automatically detect the locations of the titles, axes, and text within a graph.

An issue that we encountered in processing the eye track data was that a lot of eye locations were invalid. Either because the eye tracker had low confidence in its prediction of where a participant was looking or the participant was not looking at the screen. These errors seemed to be concentrated on certain participants. Several participants had no valid eye data for the entire study. To try and clean up the fixation data used we discarded all the eye gaze information for any participant with less than 10% valid eye gaze locations over the entire study. This rule discarded almost half of all of the participant gaze information. The issues with the eye gaze locations may be due to issues with the Gazepoint eye tracker and its ability to track eye movement when participants wore corrective lenses. This is a common issue with eye trackers due to the way certain eye trackers calculate the pose of the eye [41]. Even after removing the eye tracks of users whose eye gaze couldn't be tracked effectively the wide variation in the inter-subject consistency and low average inter-subject consistency prevented the establishment of a true upper bound in predictive accuracy of the saliency models using the AUC similarity measure though the KL divergence similarity measure was more insensitive to issues with lost eye gaze samples.

The issues with the missing eye gaze locations from the participants we did use also affected the results of the saliency model evaluation for both similarity measures, since the performance of the saliency models was much higher when the eye gaze information from the MIT1003 dataset was used for the natural images. Even though the KL divergence measure was more insensitive there was a distinct increase in the similarity measure score for the MIT1003 dataset. This suggests that with higher confidence eye gaze information the measured predictive quality of saliency models on SGs would also increase. Looking forward it would be ideal to evaluate the predicting accuracy of

saliency models using the SGs and eye gaze information from the MASSVIS dataset to see if the SG based biasing improves the performance of a standard visual saliency model [11].

References

1. Kosslyn, S.M.: Understanding charts and graphs. *Appl. Cogn. Psychol.* **3**, 185–225 (1989)
2. Harrison, A., Etienne-Cummings, R.: An entropy based ideal observer model for visual saliency. In: 2012 46th Annual Conference on Information Sciences and Systems (CISS), pp. 1–6. IEEE (2012)
3. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: *Advances in Neural Information Processing Systems*, pp. 545–552 (2007)
4. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998)
5. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: SUN: a Bayesian framework for saliency using natural statistics. *J. Vis.* **8**, 32.1–32.20 (2008)
6. Gattis, M., Holyoak, K.J.: How graphs mediate analog and symbolic representation. In: *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (1994)
7. Aumer-Ryan, P.: Visual rating system for HFES graphics: design and analysis. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 2124–2128 (2006)
8. Fausset, C.B., Rogers, W.A., Fisk, A.D.: *Visual graph display guidelines*, Atlanta, GA (2008)
9. Gillan, D.J., Wickens, C.D., Hollands, J.G., Carswell, C.M.: Guidelines for presenting quantitative data in HFES publications. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **40**, 28–41 (1998)
10. Petkosek, M.A., Moroney, W.F.: Guidelines for constructing graphs. In: *Human Factors and Ergonomics Society Annual Meeting*, pp. 1006–1010 (2004)
11. Borkin, M.A., Bylinskii, Z., Kim, N.W., Bainbridge, C.M., Yeh, C.S., Borkin, D., Pfister, H., Member, S., Oliva, A.: Beyond memorability: visualization recognition and recall. *IEEE Trans. Vis. Comput. Graph.* **22**, 519–528 (2016)
12. Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H.: What makes a visualization memorable? *IEEE Trans. Vis. Comput. Graph.* **19**, 2306–2315 (2013)
13. Vessey, I., Galletta, D.: Cognitive fit: an empirical study of information acquisition. *Inf. Syst. Res.* **2**, 63–84 (1991)
14. Vessey, I.: Cognitive fit: a theory-based analysis of the graphs versus tables literature. *Decis. Sci.* **22**, 219–240 (1991)
15. Ngo, D.C.L., Samsudin, A., Abdullah, R.: Aesthetic measures for assessing graphic screens. *J. Inf. Sci. Eng.* **16**, 97–116 (2000)
16. Zen, M., Vanderdonckt, J.: Towards an evaluation of graphical user interfaces aesthetics based on metrics (2014)
17. Acartürk, C.: Towards a systematic understanding of graphical cues in communication through statistical graphs. *J. Vis. Lang. Comput.* **25**, 76–88 (2014)
18. Greenberg, R.A.: *Graph comprehension: difficulties, individual differences, and instruction* (2014)
19. Halford, G.S., Baker, R., McCredden, J.E., Bain, J.D.: How many variables can humans process? *Psychol. Sci.* **16**, 70–76 (2005)
20. Pinker, S.: *Theory of graph comprehension* (1959)

21. Trickett, S.B., Trafton, J.G.: Toward a comprehensive model of graph comprehension: making the case for spatial cognition. In: Barker-Plummer, D., Cox, R., Swoboda, N. (eds.) *Diagrams 2006*. LNCS (LNAI), vol. 4045, pp. 286–300. Springer, Heidelberg (2006). doi:[10.1007/11783183_38](https://doi.org/10.1007/11783183_38)
22. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980)
23. Oliva, A., Torralba, A., Castelhana, M.S., Henderson, J.M.: Top-down control of visual attention in object detection. In: *Proceedings 2003 International Conference on Image Processing* (Cat. No. 03CH37429), p. I-253-6. IEEE (2003)
24. Gao, D., Vasconcelos, N.: Integrated learning of saliency, complex features, and object detectors from cluttered scenes. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, pp. 282–287. IEEE (2005)
25. Gao, D., Vasconcelos, N.: Discriminant saliency for visual recognition from cluttered scenes. *Adv. Neural. Inf. Process. Syst.* **17**, 1 (2004)
26. Torralba, A., Oliva, A., Castelhana, M.S., Henderson, J.M.: Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* **113**, 766–786 (2006)
27. Parkhurst, D.J., Law, K., Niebur, E.: Modeling the role of saliency in the allocation of overt visual attention. *Vis. Res.* **42**, 107–123 (2002)
28. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* **40**, 1489–1506 (2000)
29. Zhao, Q., Koch, C.: Learning a saliency map using fixated locations in natural scenes. *J. Vis.* **11**, 1–15 (2011)
30. Chauvin, A., Herault, J., Marendaz, C., Peyrin, C.: Natural scene perception: visual attractors and images processing. In: *Connectionist Models of Cognition and Perception - Proceedings of the Seventh Neural Computation and Psychology Workshop*, pp. 236–248. World Scientific Publishing Co. Pte. Ltd., Singapore (2002)
31. Lin, Y., Fang, B., Tang, Y.: A computational model for saliency maps by using local entropy. In: *AAAI Conference on Artificial Intelligence* (2010)
32. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* **4**, 219–227 (1985)
33. Peters, R.J., Iyer, A., Itti, L., Koch, C.: Components of bottom-up gaze allocation in natural images. *Vis. Res.* **45**, 2397–2416 (2005)
34. Kadir, T., Brady, M.: Saliency, scale and image description. *Int. J. Comput. Vis.* **45**, 83–105 (2001)
35. Tamayo, N., Traver, V.J.J.: Entropy-based saliency computation in log-polar images. In: *Proceedings of the International Conference on Computer Vision Theory and Applications*, pp. 501–506 (2008)
36. Wang, W., Wang, Y., Huang, Q., Gao, W.: Measuring visual saliency by site entropy rate. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2368–2375. IEEE (2010)
37. Bruce, N.D.B., Tsotsos, J.K.: Saliency based on information maximization. *Adv. Neural. Inf. Process. Syst.* **18**, 155–162 (2006)
38. Bruce, N.D.B., Tsotsos, J.K.: Saliency, attention, and visual search: an information theoretic approach. *J. Vis.* **9**, 5.1–5.24 (2009)
39. Itti, L., Baldi, P.: Bayesian surprise attracts human attention. *Adv. Neural Inf. Process. Syst.* **18**, 547 (2006)
40. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 2106–2113. IEEE (2009)
41. Dahlberg, J.: Eye tracking with eye glasses (2010)

42. Tatler, B.W.: The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *J. Vis.* **7**, 4 (2007)
43. Hart, B.M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., Einhäuser, W., Marius, B., Vockeroth, J., Bartl, K., Schneider, E., Einhäuser, W.: Gaze allocation in natural stimuli: comparing free exploration to head-fixed viewing conditions. *Vis. Cogn.* **17**, 1132–1158 (2009)
44. Schumann, F., Einhäuser-Treyer, W., Vockeroth, J., Bartl, K., Schneider, E., König, P.: Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *J. Vis.* **8**, 12.1–12.17 (2008)
45. Bylinskii, Z., Borkin, M.A.: Eye fixation metrics for large scale analysis of information visualizations. In: *ETVIS Workshop on Eye Tracking and Visualization* (2015)
46. Borji, A., Itti, L.: State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 185–207 (2013)
47. Cooper, R.A., Plaisted-Grant, K.C., Baron-Cohen, S., Simons, J.S.: Eye movements reveal a dissociation between memory encoding and retrieval in adults with autism. *Cognition* **159**, 127–138 (2017)
48. Tilke, J., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: *Proceedings of IEEE International Conference on Computer Vision*, pp. 2106–2113 (2009)