Eye-Tracking Study Using Cellular Automaton Patterns as Visual Stimuli: Implications for Current Models of Stimulus-Driven Selection Processes*

Anirudh Tiwathia[†]

Cognitive Science Department,
Vassar College, 124 Raymond Avenue,
Poughkeepsie, New York, USA, 12604

Cristián Opazo-Castillo

Physics and Astronomy Department,
Vassar College, 124 Raymond Avenue,
Poughkeepsie, New York, USA, 12604

This study examines goal-free viewing of cellular automaton (CA) images to address the nature of the bottom-up process, the robustness of salience as a framework for explaining fixation points, and the particular features that can characterize salience. The influence of familiarity on oculomotor strategy is also addressed. A qualitative study of the results show promising trends. Higher-level structural features such as pockets of regularity within randomness or localized structures within regularity were salient for most participants. These results raise interesting questions about the kinds of visual features that can be used to characterize salience. An unexpected result is that many fixations occur in blank regions within images featuring nested (fractal) structures. Many of these findings escape current psychophysical models of oculomotor strategy. Hence, future eye-tracking studies with CAs as stimuli could greatly improve our current understandings of the human visual system.

1. Introduction

1.1 Perception: Limiting and selecting information

The human experience of the external world is enabled by the various methods of perception and analysis at our disposal. Thus to understand

---

†Present address: Apt 412, The Manor, 333 E 43rd Street, New York, NY, 10017
Electronic mail address: anirudh.tiwathia@gmail.com.

the human experience we must examine perception and determine the ways in which it affects the information we receive.

Generally speaking, perception can be thought of as a mechanism that limits and selects the information received by the agent from its environment. Due to the design of the instruments of perception, only a section of the global array of information is sampled and then only a segment of the information sampled is further analyzed as necessary.

Nowhere is the phenomenon of limiting and selecting information more evident than in the visual system. For example, the human eye is composed of only three kinds of color-sensitive cells—one each for red, green, and blue. Thus only three regions of the light spectrum are sampled while the rest are ignored [1]. Similarly, the amount of information coming down the optic nerve at any point is estimated to be on the order of 108 bits per second—this much information is far too large for the brain to process and assimilate into the conscious experience [2]. Moreover, human eyes do not have a uniform visual response. The best visual acuity is limited to the central 5° of the visual field (known as *foveal vision*). For nonfoveal parts of the visual field we must rely on cruder representations, with acuity sharply falling off from the center. This loss in acuity is mainly due to decreasing retinal cell density and increases in receptive-field center size as one moves away from the retinal center [3].

Thus, contrary to common belief, we do not see a majority of things around us. As the study of inattentional blindness demonstrates, even salient or distinctive objects in clear sight will not necessarily be "seen" by an observer attending to another object in the scene. Though the observer will have no trouble noticing the object once it is pointed out [4].

1.2 Perception: Mechanisms of selection

Since only a sliver of the available visual information is processed, there must be efficient mechanisms in place to select the behaviorally relevant information. Two attentional selection mechanisms are thought to control this process: bottom-up and top-down.

Bottom-up attentional selection is characterized by fast, saliency-based, stimulus-driven selection mechanisms. Support for the existence of bottom-up mechanisms comes from data demonstrating that visual attention can be captured under the right stimulus condition, for example, highly salient feature singletons, such as a red symbol among a group of blue symbols, will immediately capture attention [5].

Top-down attentional selection is characterized by slower, goal-directed selection mechanisms that are governed by the observer's expectations or intentions. The simplest evidence for top-down mechanisms is the ability of an observer to volitionally allocate attention to specific

objects or regions of space. Further evidence of top-down mechanisms comes from the effect of task instructions on both patterns of fixation and spatiotemporal dynamics of eye-movements [5].

The relationship between these two attentional selection mechanisms is poorly understood. A significant portion of the research addressing this issue examines eye-movement patterns of observers viewing different scenes under different task conditions.

1.3 Methodology: Eye-tracking and attention

Eye-tracking is considered a valid method for studying visual attentional mechanisms because it has been shown that eye-movement and attention are generally related. Intuitively, this assumption seems reasonable given that both eye-movements and attention are involved in selecting the most relevant region in the visual field. Moreover, even though the attentional focus and fixation point can be at separate locations, there exists psychophysical evidence suggesting that focal attention at the site of pending eye-movement is a necessary precursor for the completion of that eye-movement [5]. Also, brain studies with monkeys have shown that shifts in attention during different tasks are associated with eye-movement preparations [6]. Other electrophysiological data also indicate that the two mechanisms share some of the same neurophysiology [7].

1.4 Methodology: Data collection with eye-tracking

Early work in eye-movement analysis showed that, when viewing a visual scene, observers do not scan the scene, but instead fixate on certain locations and move between them.

Currently it is possible to categorized eye-movements into the following: (a) saccadic—voluntary rapid eye-movement from one point to another, (b) miniature—group of involuntary eye-movements that cause the eyes to waver, (c) pursuit—smooth involuntary eye-movement to keep a moving object foveated, (d) compensatory—similar to pursuit but the head moves while fixation is maintained, (e) vergence—relating to the focusing on a near or far object, (f) optokinetic—involuntary zigzag movement performed when observing repeated moving patterns [8].

Most studies concerning eye-movements concentrate on saccadic movement and fixations. The main purpose of a saccade is to foveate a new region of interest so as to get higher-resolution input. After a saccade moves the gaze of the eye to a new region, the eye must dwell on this new region for at least 100 ms so as to minimize image blur and allow the visual system to process the new information. This point is known as a *fixation*. For each fixation, an area around the point of fixation, the size of which depends on stimulus density, can be successfully processed [9].

1.5 Models of oculomotor strategies

The factors that govern where fixations will occur in a given image are still controversial. Some studies claim that fixation location is random, while others claim that stimulus factors are critical, and yet others place emphasis on cognitive factors. Studies showing consistency in fixation locations for different observers demonstrate that fixation locations are not random, and that the human visual system is designed in such a way that competent observers naturally utilize similar oculomotor strategies [7].

As with most aspects of cognitive science, constructing a model of oculomotor strategies has proven quite difficult. Most models of oculomotor strategies have emerged from laboratory-based studies of visual search. These various models differ in their emphasis on bottom-up (stimulus based) and top-down (task based) factors.

In the 1980s, Triesman and Gelade introduced the feature integration theory (FIT) as a bottom-up model for visual search. The FIT consisted of two stages: a parallel and a serial search. In the parallel search, the target object was different enough from the distractors that all objects could be processed in parallel to identify the target. In such a situation, the time taken to perform the search would be largely independent of the number of objects in the search space (set size), as cited in [8]. For example, a red 7 among a collection of blue 7s could be found with a parallel search. On the other hand, the serial search occurs when the target object is not fundamentally different from the distractors, for example, a randomly oriented L among a group of randomly oriented Ts. In this case, each item has to be considered and a decision must be made whether it is a target or not—for a serial search the search time is directly correlated to the set size, as cited in [8].

In the FIT model, the presented scene is processed in parallel for each individual basic feature (color, orientation, etc.) and then the scene is encoded in feature maps. Feature maps encode the activity in response to certain features whereas a "master map" encodes information about location, spatial distribution within a feature map, and the relationships between different feature maps. A parallel search is performed when only a single basic feature is necessary for detection; whereas a serial search is performed when relationship between features is necessary, as cited in [8].

A similar bottom-up oculomotor model, the visual salience model, suggests that people direct their gaze at the most visually salient location in the retinal image. There is significant evidence to suggest that preattentive, parallel levels of processing do not represent all parts of a visual scene equally, but instead that a weighted representation is derived with certain parts eliciting stronger responses than others [2]. Whether a given part of the scene elicits a strong response is thought to be depen-

dent on the "context" in which it is presented, that is, it is dependent on what other stimuli are present in the other parts of the visual field [2].

The visual salience model claims that to accomplish preattentive selection, an explicit two-dimensional map (known as the *saliency map*) encodes the saliency of objects in the visual environment. The saliency map is similar to the master map in FIT, which is derived from the integration of various feature maps. Competition among neurons within the saliency map gives rise to a single winning location that is considered the most salient location and chosen as the next target for saccade and fixation. If this location is inhibited due to prior fixation then the next most salient location is chosen and so on [2].

Many computational models have in some form or another accepted the idea of a saliency map to compute the next gaze location. Electrophysiological evidence also points to the existence of several neuronal maps that appear to encode for saliency of a visual stimuli [2]. However, some researchers oppose the idea of a topographic map in the brain whose specific purpose is to represent salient stimuli. Desimone and Duncan have suggested that selective attention results from the interaction between feature maps, each of which implicitly encodes the saliency of a stimulus with respect to one given feature, as cited in [2]. Although it remains debated whether or not saliency is expressed explicitly in one or more visual field maps, most bottom-up models of visual search rely on some manifestation of a saliency-based decision making process.

Unlike the FIT and the visual salience model, the guided search oculomotor strategy includes a top-down approach. Here, the fundamental idea is that information about the nature of the target can bestow greater importance to specific features and as a result bias the direction of gaze. Within the guided search model, there is an initial stage of parallel processing in which the visual field is evaluated with regard to similarity to the expected target. The items that are considered similar enough to the target are then passed onto a serial phase where they are individually inspected in greater detail. Thus, gaze is directed to the item with the greatest similarity to the expected target, after which gaze is successively shifted to items of decreasing similarity [7]. Experiments have shown that observers can use features such as target color and shape to guide attention or gaze in visual search. Thus the search is only performed in a preselected subgroup [10]. Moreover, eye-movement data from a real-world task, such as tea making and sandwich preparation, provide strong evidence that, in such cases, the eyes are driven primarily by top-down information and rather little by the intrinsic salience of objects [11]. That being said, no experiments thus far have been shown to rule out saliency maps as the mechanisms for selecting and prioritizing locations for further processing.

1.6 Competing influences: Top-down and bottom-up

Parkhurst *et al.* performed a study examining the extent to which bottom-up factors play a role in allocating attention by comparing actual fixation locations to computed stimulus salience as determined by a biologically motivated, saliency-based, bottom-up model. They attained results that indicate the presence of stimulus-driven mechanisms even under natural viewing conditions where top-down mechanisms would be active [5].

Interestingly, Parkhurst *et al.* also found a greater correlation between calculated stimulus salience and measured fixation locations for fractal images than for other kinds of images [5]. They put forth two possible explanations for why fractal images might give rise to a greater correlation between calculated salience values and measured fixation points.

Their first explanation suggests that the increased correlation for fractal images could be caused by differences in stimulus characteristics. They found that there was often a greater separation between the salience of peaks and the average background level, that is, there were fewer areas of high salience and the difference in salient points and background levels was higher than for other images [5]. Thus, the sharp contrasts in fractal images with regard to certain features may increase the influence for stimulus-driven mechanisms.

The second explanation motivates top-down influences by suggesting that the eye movements could have been influenced by previously developed attentional biases. For example, they found that, given images of home interiors, participants often scanned tabletops independent of their salience in the presented image. This behavior seems quite reasonable since scanning tabletops is most often the best place to find objects of interest. Moreover, they argue that the influence of such top-down attentional biases should give rise to greater inter-participant variability for fixation location than the influence of stimulus-driven attentional allocation. Their data show exactly this result. The inter-participant variability for fixation locations was lower for fractals than all other image types. This result agreeing with their earlier assertion that bottom-up mechanisms play a greater role in fractal images [5].

1.7 Role of familiarity

In the Parkhurst *et al.* study, the difference in degree of top-down attentional bias between different image types could have resulted from differences in levels of exposure. Besides fractal images, the other image categories in the study were: home interiors, buildings and city scenes, and natural landscapes. The exposure of most participants to scenes that fall in those three categories is probably greater than their exposure to fractal images. However, the effect of familiarity with visual stimuli on oculomotor strategy has not been fully disambiguated.

Some studies do suggest that visual salience may be affected by repetition of stimuli [12] or by familiarity with stimuli [13]. In fact, some studies have found results suggesting that if a global representation (e.g., a line-figure of a face) is available then lower level representations of the constituent parts (e.g., curvatures of lines) can possibly be suppressed even if it would be quicker to process the stimulus-set using the low-level features [13]. These results are consistent with the assertion that limited exposure to fractal images could allow for greater reliance on low-level, stimulus-driven oculomotor strategies. In fact, it seems commonsensical that extended exposure to a stimulus would facilitate the development of top-down, attentional biases towards that stimulus.

▌ 1.8 Present study: Examining bottom-up processes

The present experiment aims at examining the nature of bottom-up mechanisms in the human visual system using the eye-tracking paradigm. To achieve this goal, a two-pronged approach was taken to minimize the influence of top-down mechanisms, and thereby allow for a better examination of the bottom-up mechanisms at work.

First, the task-instructions explicitly specified that all participants were to engage in "spontaneous looking," without any specific task, target, or goal in mind. This reduced observer expectations that could have been used by top-down mechanisms to determine a subset of fixation locations within the presented stimuli.

Second, the stimuli chosen in this study were two-dimensional black-and-white abstract images generated by taking snapshots of the time evolution of cellular automaton (CA) systems. This choice of stimuli held numerous advantages in the context of reducing top-down influences.

Images generated by CAs have limited exposure among the general populace and thus it is unlikely for participants to have previously developed attentional biases. Also, although they possess structural features found in nature such as repetition, nesting, and randomness, CA-generated images are quite abstract in comparison to natural or representational scenes. It agrees with intuition that abstract images, devoid of familiar features, would reduce the influence of top-down selection. The data about fractal images presented by Parkhurst *et al.*, as discussed earlier, supports such an assertion. It thus seems reasonable to believe that the use of CA images would encourage greater influence of bottom-up mechanisms on fixation patterns.

Moreover, the use of such images also allowed us to examine whether, given the minimization of other top-down influences, higher-level structural features (e.g., nesting, randomness, or repetition) would directly influence patterns of fixation.

Visual studies concerning CA images are also of interest to the newly developing field of pure New Kind Of Science (NKS). Stephen Wolfram, in his descriptions of the methodologies of NKS, places significant importance in the ability of the human visual system to analyze, within reasonable limitations, the outputs of different CAs [1]. It would be useful to determine whether inexperienced eyes instinctively fixate on features most commonly studied in CA images and also to examine, how—if so—previous exposure to CA-generated images changes fixation patterns.

In summary, the goals of this study are: (*a*) to encourage and subsequently examine bottom-up mechanisms involved in guiding gaze-direction and selection of fixation points; (*b*) to examine the ways in which participants observe different CA images to notice if instinctive fixation points coincide with those features often studied in the field of NKS; (*c*) to determine if structural features such as nesting, repetition, and randomness specifically influence oculomotor strategy; (*d*) to examine the effect of past experience with CA images and CA theory on fixation locations so as to comment on the relationship between familiarity and oculomotor strategy.

2. Methods

2.1 Participants

A total of 18 participants were tested: nine participants were naïve to the theory of CAs, NKS, and had little or no exposure to images generated by CAs; whereas the other nine participants were students of the NKS Summer School and had significant exposure to complex systems theory, CA formalism, and NKS.

All participants had normal or corrected to normal vision, with no handicaps such as dyslexia or color-blindness. All participants had little or no prior exposure to an eye-tracking device or to theories resulting from research conducted using an eye-tracking device.

2.2 Stimulus

All the stimuli presented to the participants were images generated by one-dimensional (1-D) CAs. The CA images were generated by selecting an initial condition (a 1-D array of ones and zeroes depicted as black and white cells) and then applying a simple rule to the initial condition for a specific number of iterations. Thus each two-dimensional (2-D) CA image actually represented the evolution of a 1-D CA over time—any horizontal line of the CA image at any given step represents the state of the 1-D CA at that point in time (see Figure 1).

The initial conditions chosen were either a row of white cells with a single black cell at the center, or a random distribution of black and
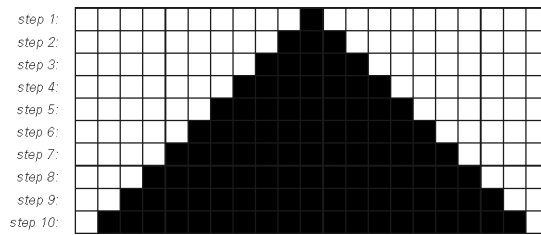
Figure 1. Evolution of a simple CA. Here is a visual representation of the behavior of a CA, with each row of cells corresponding to one time step. At the first step the cell in the center is black and all other cells are white. Then on each successive step, a particular cell is made black whenever it or either of its neighbors was black on the step before. As the picture shows, this leads to a simple expanding pattern uniformly filled with black. (Image Source: S. Wolfram, *A New Kind of Science*.)

white cells generated using *Mathematica*®. The CA images were generated by applying a specific two-neighbor rule for either 600 (for isosceles triangular patterns from simple initial conditions), 900 (for other patterns from simple initial conditions), or 1024 (for patterns from random initial conditions) iterations. All CA images were generated with *Mathematica*. It was ensured that there was at least one pixel per cell of CA image. Each image was approximately 53×53 degrees of visual angle in size when viewed from a distance of 18 inches.

Wolfram has classified CA patterns into four different categories: *simple*, *nested*, *random*, and *localized structures*. The simple CA images featured either a uniform pattern (e.g., an all black structure; see Figure 1) or a repetitive pattern (e.g., black and white checkerboard). The nested images featured fractal structures such that the image contained replicates of itself (see Figure 2). The random patterns showed no overall regularities or localized structures (see Figure 3). Patterns with localized structures showed both regions of regularity and regions of randomness but also showed structures that propagated across the time evolution [1].

The method for enumerating rules in this study is consistent with that outlined by Wolfram in [1]. All images were composed of only black, white, or gray cells.

2.3 Design

Participants were seated at a testing table and were presented a series of CA images on a 19 inch flat-screen color monitor. While at the testing table, the participants were seated at the eye-tracking device that was attached to the table. With their head in the eye-tracking "head-positioner" the participants' eyes were approximately 18 inches
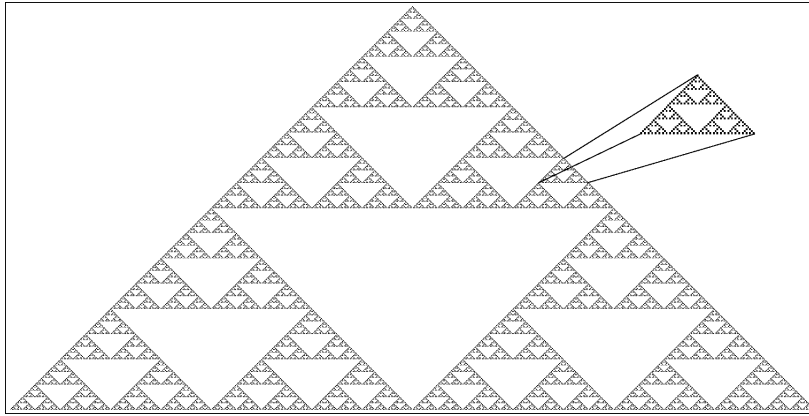
Figure 2. A CA that generates a nested pattern. Each triangular section is essentially just a smaller copy of the whole pattern, with still smaller copies nested inside it. Patterns with nested structure of this kind are often called "fractal" or "self-similar." (Image Source: S. Wolfram, *A New Kind of Science*.)
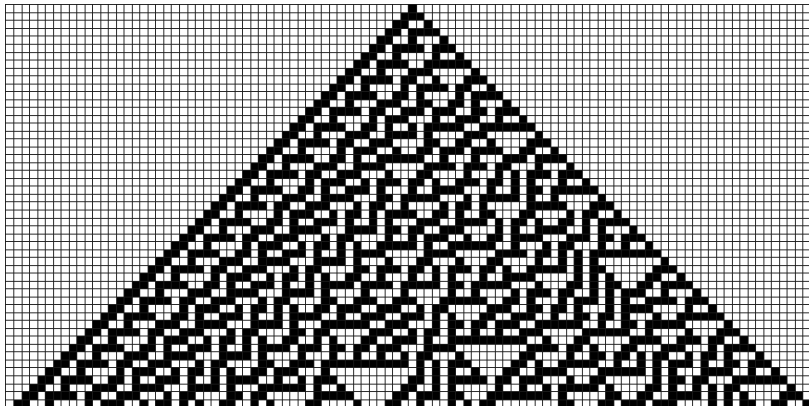
Figure 3. A CA that generates a random pattern. This particular CA is known as Rule 30. The transformation rule is shown at the bottom. (Image Source: S. Wolfram, *A New Kind of Science*.)

away from the computer screen. When looking straight ahead, the participant's gaze was horizontally centered and at approximately two-thirds of the screen height from the bottom of the screen.

All images presented were centered on the screen; when centered, all images were large enough to leave at most an inch boundary from the edge of the screen. The stimuli presented were as follows (in order): four 2-D arrays of CA image thumbnails, 20 individual CA images with simple initial conditions, and nine individual CA images with random initial conditions. The four 2-D arrays were presented for 10 seconds each, both the individual CAs from simple and random initial conditions were presented for 8.5 seconds each.

Each 2-D array had 20 thumbnails of CA images (five rows, four columns). Out of the 20 individual CA images from simple initial conditions: five were simple, five were nested, six were random, and four were considered to have localized structures. Out of the nine individual CA images from random initial conditions: one was simple, one was nested, three were random, and four were considered to have localized structures.

Within each image set (array of images, simple initial conditions, random initial conditions) the distribution of images from the four CA categories was performed randomly. The sequence of images was kept constant for all participants (for concerns regarding order effects, see section 5.1).

Calibration of the eye-tracker was performed before the start of the experiment, and then a second time after the 20 CA images from simple initial conditions had been presented.

The eye-movement data were collected on a monocular "ViewPoint Eye-Tracker" designed by Arrington Research, Inc. The IR camera always focused on the right eye of the participant. All stimuli were displayed using the ViewPoint PC-60 (version 2.8.1) software. The data were collected using default settings with low resolution. The eye-tracker sampled eye position every 20 milliseconds.

2.4 Participant instructions

The participants were informed that they would be presented patterns generated by CAs and they were also informed of the order of the category of images (arrays of CA images first, CA images from simple initial conditions next, and finally CA images generated from random initial conditions). They were told that some of the patterns were relatively simple while others were more complicated. Participants were instructed to observe each image as they saw fit. In order to remove expectation of later questions, they were explicitly told that they would be neither asked to report nor explain any of their observations. It was emphasized that the observations did not entail any kind of further analysis on

the part of the participant. This was done to remove any expectations of a specific task or goal so as to bring the participant into a state of "spontaneous looking" and thereby reduce top-down influences.

The participants were also warned about the importance of minimizing head-movements so as to prevent errors in data collection.

2.5 Procedure

Each participant was asked to view all three sets of images. After receiving instructions, the participants were seated at the eye-tracker. After calibration, they were presented the sequence of images and their eye-movement data was recorded for future analysis. There was no communication with the participant for the duration of the image presentations. While the experiment was in progress, the researcher monitored the output from the eye-tracker on a secondary screen so as to enable greater accuracy in data collection. The participants were offered the option to take a break before the second calibration, which was declined by all of them.

2.6 Data analysis

Due to the low number of participants and the particular constraints of our experiment, the data was not subjected to standard statistical analysis; instead, the paths of eye-movements were analyzed qualitatively. This was considered an adequate method of analysis since this study aimed to discover interesting phenomena and motivate additional study of these phenomena. We assert that any truly meaningful results should be noticeable by careful qualitative visual analysis of oculomotor data.

3. Results

The trends mentioned here are meant to motivate future studies where statistical analysis of eye-movement data would be of greater use.

3.1 Simple or repetitive cellular automaton images

For simple or repetitive CA images, no distinct common points of fixation were observed. There was significant inter-participant variability with regard to oculomotor strategy and the patterns of fixation seemed quite arbitrary.

3.2 Images with pockets of regularity

CA patterns that appear to be mostly random displayed no distinct common points of fixation within the regions of randomness. However, in the cases where random patterns had pockets of regularity, a majority of participants did fixate on the regions that displayed regularity. Regions
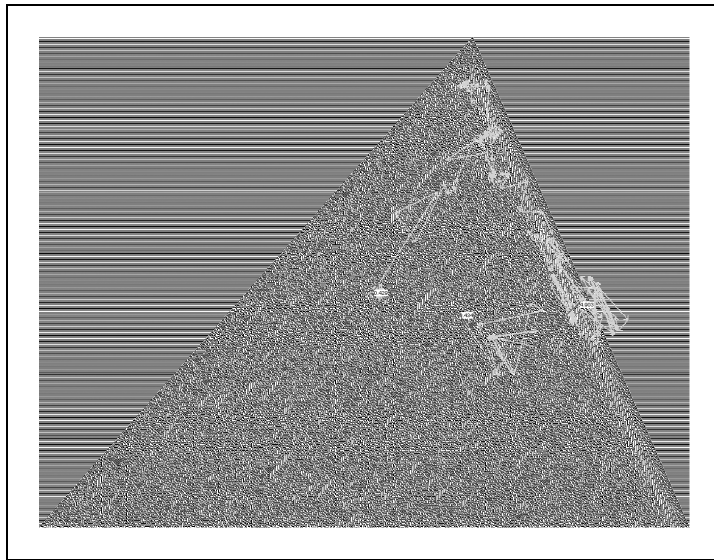
Figure 4. Eye-movement data when viewing Rule 101 from simple initial conditions. Though this rule shows random behavior in most regions, it also possesses regions of regularity (right-hand boundary of triangle). This region enjoyed fixation from the majority of participants—thereby raising questions about the kind of higher-level features (e.g., regularity) that can be used by the visual system to determine salience.

of regularity were defined as regions that showed an obvious underlying order. This behavior was noticed in eye-tracking data collected from CA images generated by rules such as 30, 101, and 73 (see Figure 4).

3.3 Images with localized structures

For CA patterns classified as possessing localized structures, most participants demonstrated fixation patterns that loosely followed these localized structures. However, the selection of particular localized structures to examine, the order of selection, and the specific manner in which the localized structures were examined all showed great inter-participant variability and thus did not suggest any obvious common oculomotor strategy.

3.4 Images with nested structure

The data collected from nested CA images also provides some interesting results. To elucidate these results it is useful to separate the nested images into two subgroups.
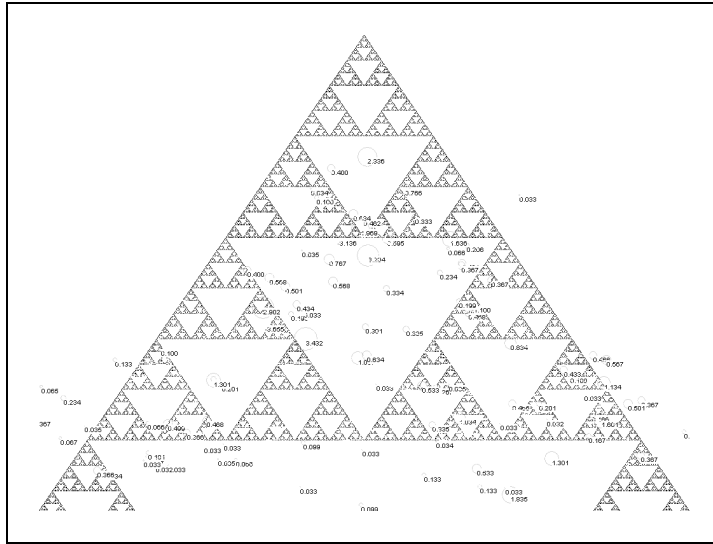
Figure 5(a). Data for all nine naïve participants when viewing Rule 60. Eye-tracking data from rules such as 60 showed that the majority of participants tended to fixate on the blank regions within the triangles.

In the first subgroup, rules such as 60, 90, and totalistic Rule 948, were categorized together strictly on the basis of nested structure. Within this first subgroup, eye-movement data showed that for most participants a considerable number of fixations were in the large blank spaces (see Figures 5(a) and 5(b)).

The nested CA images rules such as 105 and 150 were categorized as a separate subgroup because they both share the following properties: nested structure, symmetry, and an explicit vertical axis of symmetry. Though both Rules 90 and 948 show symmetry, they do not have an explicit axis of symmetry. In this second subgroup (Rules 105 and 150), eye-movement data fell into two broad categories: in one case participants tended to pick one side (left or right of the symmetry axis) and constrain their visual search to this region exclusively (see Figure 6(a)). Some participants made short excursions to the other side but returned quickly to the original side. In the other case, participants tended to show quick, localized eye-movements that were concentrated on the region around the symmetry axis and its immediate neighborhood (see Figure 6(b)).

3.5 Effect of familiarity

For the CA image generated by Rule 110 (categorized with images possessing localized structures) the majority of participants without prior experience with CA images fixated on the "glider" at the right-hand
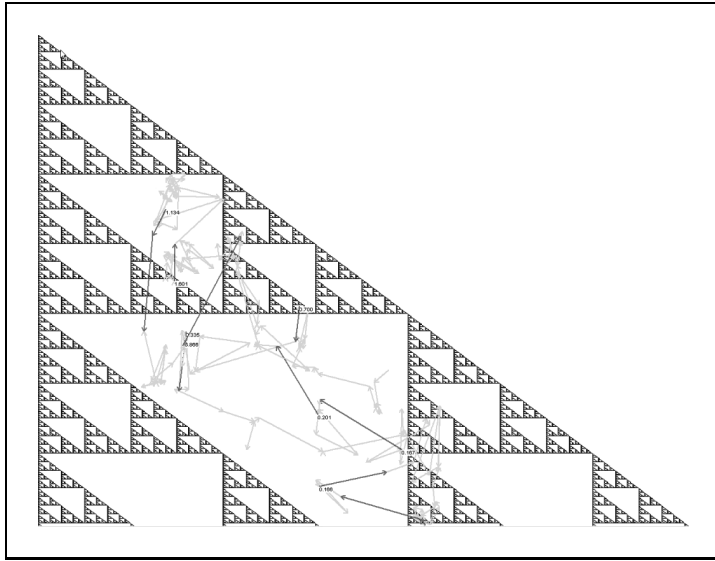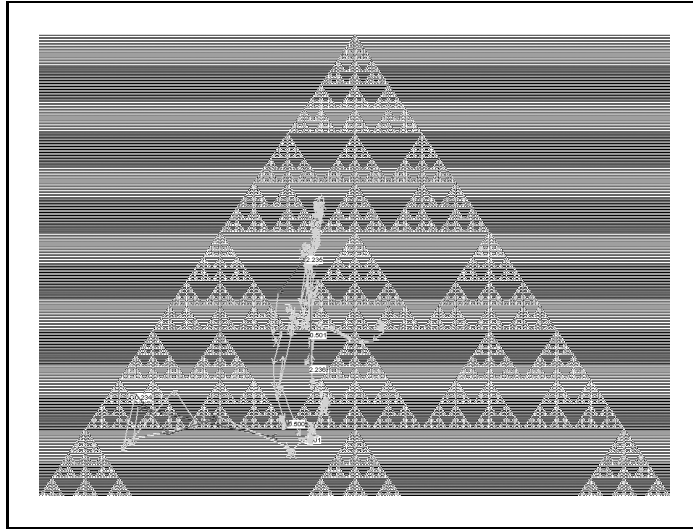
Figure 5(b). Eye-movement data from participant viewing nested CA. This image shows the eye-movement path for one experienced participant. Both experienced and naïve participants tended to fixate in these blank regions.

side of the image (see Figure 7(a)). Interestingly, this and other gliders in Rule 110 are often the subject of attention and study. On the other hand, none of the participants that had prior experience with CA images fixated on this glider, thereby providing a clear example of differences in oculomotor strategies between these groups (see Figure 7(b)). It should be noted here that there was little or no apparent difference between the experienced and naïve groups with regard to the above-mentioned behavioral tendencies.

Another difference between these two groups related to *clustering effects*. It was observed that the eye-movement patterns of the experienced group as a whole were more clustered together than the eye-movement patterns of the naïve group as a whole (see Figure 8). In other words, there was greater spatial inter-participant variability among the naïve group than the experienced group. Also, for many CA images, not only was there a clustering effect but also the cluster of eye-paths was higher in the screen. Thus, in these images, the eye-paths of the experienced were more concentrated in regions displaying earlier time steps.

It should be noted that the clustering effect was found for all individual CA images across all rules and initial conditions. However, no clustering difference was observed for the four images that contained arrays of images generated by CAs (Figure 9). The clustering effect was only present in images with single CA evolutions.

Figure 6(a). Participant showing side-picking behavior. This rule generates a pattern which is nested, symmetric, and has an explicit line of reflection. The eye-movement data from this rule, as well as from Rule 150, showed two kinds of behavior: affinity for the center column region and side-picking.
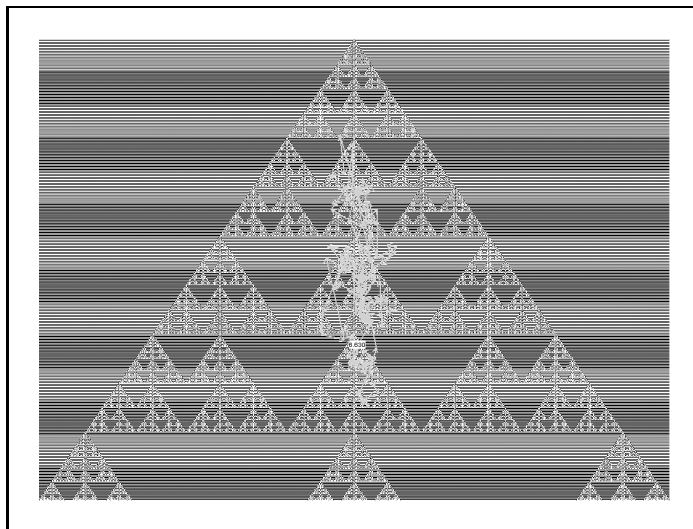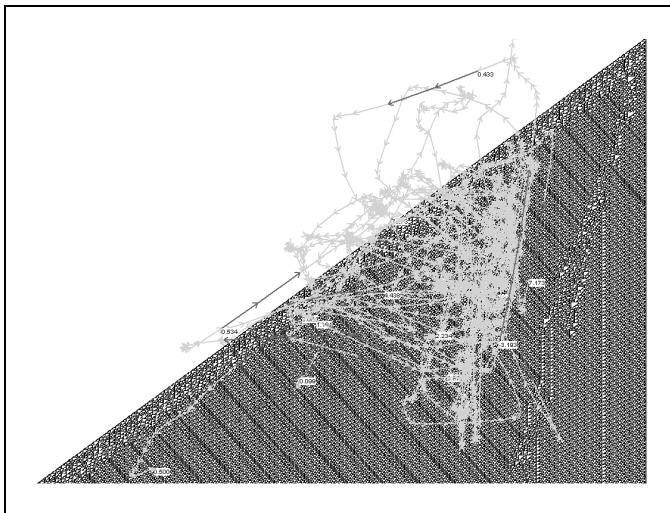

Figure 6(b). Participant showing affinity for the center column when viewing Rule 105.
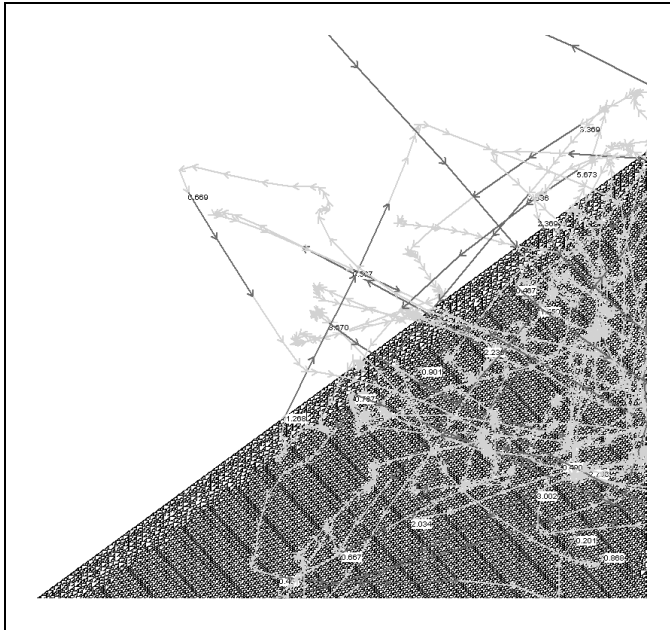
Figure 7(a). Data from all naïve participants viewing Rule 110. The fact that, unlike the majority of naïve participants, experienced participants did not fixate on the glider in Rule 110 suggests that experience has an effect on salience and oculomotor strategies.

Figure 7(b). Data from all experienced observers viewing Rule 110. It is interesting to note that none of the experienced observers fixated on the glider on the right-hand side.
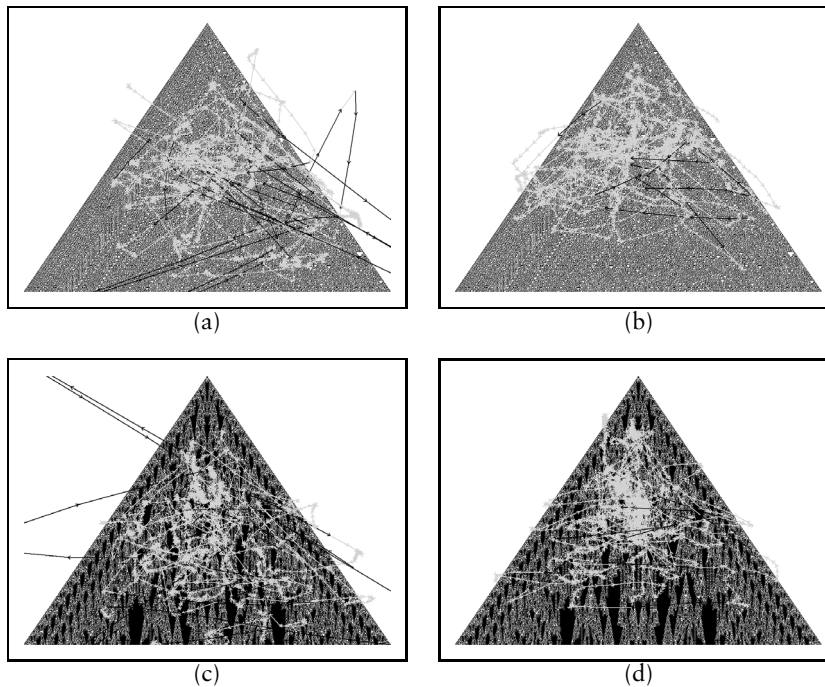
Figure 8. A strong clustering effect was seen in the eye-movement patterns within the experienced group. This decrease in inter-participant variability was hypothesized to be caused by the use of similar oculomotor strategies, which arose as a result of experience and knowledge about CA images, by experienced participants. Clustering of eye-movement paths for experienced participants was also towards the top of the image. It is suggested that experienced participants generally tended to concentrate at a higher location in the image because of top-down influences that result from their understanding that the CA image is a time evolution of a 1-D CA. (a) Rule 30 as observed by naïve participants. (b) Rule 30 as observed by experienced participants. (c) Rule T1635 as observed by naïve participants. (d) Rule T1635 as observed by experienced participants.

4. Discussion

How and to what extent do characteristics of the stimuli determine the nature of eye movement and the patterns of fixation? In this study, various steps were taken to limit the introduction of top-down information that could have affected the pattern of fixations and eye-movement in general. The eye-movement data gathered from participants engaged in task-free observations of various CA-generated images and the subsequent analysis of fixation patterns provide some interesting insight into human visual perception.

4.1 Simple patterns

No common oculomotor strategies were discovered for simple or purely random CA images. This result agrees with predictions from the visual salience model [2, 5]. In such images, especially in the absence of a particular goal, there is no means for a saliency map to consistently determine a single, unique location that corresponds to the next most salient area. In such a case, the exact location of the next point of fixation would probably be influenced by external, top-down factors including previous attentional or oculomotor strategies, boredom, or the diversion of attention to an inner thought. Such top-down influences would explain the greater inter-participant variability.

4.2 Pockets of regularity in randomness

For random CA images that have pockets of regularity, the regions of regularity were common points of fixation for many participants. The exact reason for this result is unclear, and two hypotheses emerge as competing explanations.

In such images, no lower-level features can be used to determine unique fixation locations in order of salience. It is possible that, in response to the failure of lower-level stimulus features, salience could be determined with other higher-level stimulus features such that small structures of regularity are compared with other regions of randomness. In such a case, the pockets of regularity would be highly salient in an image consisting mostly of regions of randomness.

On the other hand, a higher cognitive realization that randomness is a dominant theme in such images could result in the top-down patronage of regions of regularity as areas of interest and hence points of fixation. A systematic examination of the exact relationship between higher-level stimulus features, such as regularity or randomness, and the mechanism for determining saliency is necessary. As of now, current computational models cannot account for such a result. The subsequent investigation of this phenomenon will be necessary to allow us to update our models to account for such behavior.

4.3 Nested structures and economical fixations

The eye-movement data from nested CA images would also benefit from the examination of higher-level stimulus features as possible criteria for determining saliency. Within the first subgroup (Rules 60, 90, and 948), many points of fixation occurred on the blank regions inside the nested CA patterns. Fixations on blank areas are not a commonly reported phenomenon in eye-tracking studies. In fact, some studies explicitly reported the absence of such fixations [9].

While this study found fixations on blank regions outside the CA pattern to be extremely rare, fixations on blank regions within the CA pattern were common for the nested CA patterns. This difference in fixation frequency for blank regions inside and outside the CA pattern demonstrates that the visual system distinguishes between blank spaces within and outside of the outermost fractal structure. That being said, it is quite unclear why there were such a large number of fixations on blank areas inside the CA pattern.

It is reasonable to assume that low-level features are not guiding the pattern of fixation towards the blank areas. Since this behavior is seen for Rule 60, which is not symmetric, it is also plausible to ignore symmetry as a potential cause.

Two hypotheses are discussed below as explanations for the unexpected fixation patterns.

One possible explanation is the presence of large, empty geometric figures. Processing the geometric figure as a whole might lead the visual system to search central fixations. The central point lies within the blank regions for each nested geometric structure.

Since this behavior is seen to some degree in all nested CA images presented, another possible explanation is that the nested structure of the image is taken into account by the visual system. One could argue that, due to the nested nature of the image, by fixating inside the blank areas the observer attains a strategic position that allows for a quick evaluation of the information contained in the image as a whole.

This claim is quite reasonable considering earlier work. For any given fixation point, there is an area surrounding the fixation point within which information can be successfully extracted. This area around the fixation point can be referred to as the *conspicuity* area [14].

Motter and Belky suggest that all stimuli present in the conspicuity area are processed in parallel along with the information at the fixation point. They claim that focal attention is spread as widely as possible in the conspicuity area and that the conspicuity area can be thought of as the maximum region of focal attention for any given stimuli condition. Their study also asserts that such a view of conspicuity area agrees with earlier research [14]. Moreover, their study provides evidence that the size of the conspicuity area depends strongly on the stimulus density surrounding the point of fixation and that in general the conspicuity area has a radius of approximately twice the average nearest neighbor distance (minimum inter-stimulus distance) [14].

By fixating in the center of the elementary nested structures, participants might be gaining large conspicuity areas that allow them to quickly process the information present in that section in parallel. If this is actually what is happening then the mechanism guiding the pattern of fixation would have to be able to determine which locations are most economical with regards to conspicuity area. Although there is

reason to believe that visual engagement is extremely economical, no evidence clearly demonstrates how or what mechanisms guide fixation patterns to locations that are most economical [11]. There is also no clear evidence that shows conspicuity area to be a central criterion for determining what is economical.

Further studies with nested images should be conducted to determine how much, if any, influence the nested structure has on the patterns of fixation. At this point, however, no sound explanation can be put forth.

As a side point, it would be quite interesting to compare the fixation locations observed in this study for images such as Rule 90 with predicted fixation locations generated by current computational models of oculomotor strategy. Given the current description of their behavior, a revision of computational models should be necessary to account for fixation on blank locations.

The second subgroup of nested CA images (Rules 105 and 150) also gave rise to two other patterns of fixation. The first fixation pattern was an affinity for the region around the axis of symmetry, and the second was the concentration of fixation points on one selected side of the axis of symmetry. Both patterns are discussed below within the context of an economical visual system.

The affinity for the region around the axis of symmetry is attributed to a combination of two stimuli characteristics: symmetry with an explicit mirror (axis of reflection), and nested structures down the center column. It is hypothesized that the observer stays in the center region due to economical considerations. By fixating on just that region alone, for such images, the observer can obtain all the information about the image with the fewest eye-movements.

Similarly, economical considerations could cause the participants to pick and stick to just one side of the CA pattern. By fixation on just one half of the pattern, the observer can obtain all the information about the pattern with minimal eye-movement.

It is important to note that just symmetry in itself could not be responsible for this behavior since it is not exhibited in other symmetric patterns that do not have an explicit line of reflection. The explicit line must play some role in the recognition of the symmetric characteristics of the pattern. One could argue that simply the presence of an explicit line that separates the two halves of the image could also be responsible for influencing participant's fixations to stay on one side of the symmetry axis. However, this would not explain the affinity for the center column.

It seems reasonable that observers stay on one side of the image because of an implicit realization that, due to the evident symmetry of the image, all the information about it is available on one side; and that the presence of an explicit line of reflection aids that implicit realization. Similarly, by staying at the center column, observers can also gain all the information about the image with the fewest eye-movements.

Future study should be conducted to determine the degree to which the economical constraints of the visual system influence oculomotor strategy. Also, it must be determined what factors are taken into consideration when making economical choices. Moreover, the economical constraints on oculomotor strategy should be included in our current computational models of the human visual system.

4.4 Familiarity and fixation patterns

Besides stimulus-based results, the results also showed evidence for differences in fixation patterns between the naïve group of participants and those tested from the NKS Summer School 2004. Two main differences were found: frequency of fixation on the glider in Rule 110, and clustering.

The case of the glider in Rule 110 is a striking example of the potential differences between naïve observers and experienced ones. The majority of the naïve group found the glider to be salient. Interestingly, that glider in Rule 110 is in a sense the inverse of the pocket of regularity in random images, discussed earlier. The glider is conspicuous because its irregularity contrasts against a regular background. It seems reasonable that the mechanisms that find randomness salient among regularity are related to the mechanisms, discussed earlier, that find regularity salient among randomness.

Surprisingly, not a single experienced participant fixated on the glider. This result agrees with previous findings that suggest that visual salience can be affected by stimulus familiarity [5, 13]. This disinterest could possibly arise due to an over-exposure of the NKS Summer School 2004 participants to that particular structure. The over-exposure could increase the calculated salience for other less characteristic structures.

At this point, it is possible to ask why a similar effect was not seen among experienced participants with regard to pockets of regularity in images featuring random patterns. Two explanations are hypothesized. First, the exposure to and emphasis on pockets of regularity is significantly lower than for the glider in Rule 110. Second, asymmetries in search tasks have been demonstrated. When targets and nontargets are exchanged the nature of visual search performance also changes (e.g., it is easier to search for curved lines among many straight distractors but it is much harder to search for a straight line among many curved distractors) [2]. Similarly, pockets of randomness on a regular background are not necessarily as salient as pockets of regularity on a random background.

It still remains open to question whether pockets of regularity among randomness or pockets of randomness among regularity is more salient— however, the complete disregard for the glider by the experienced participants hints that patches of regularity among randomness might be

considered more salient and hence less susceptible to top-down inhibition.

The general difference between the oculomotor strategies of naïve and experienced participants was witnessed in the clustering effect found for many CA images. This clustering effect could be caused by many potential factors.

First of all, the NKS Summer School 2004 participants were from similar academic backgrounds as opposed to the large variety found among the naïve participants. The similar academic backgrounds could lead to the development of similar general and oculomotor strategies (there is no reason to believe this to be a strong effect). Moreover, considering the large variety in their nationalities and cultures, the common academic backgrounds among the NKS participants probably has a limited influence on the development of oculomotor strategies.

Second, the NKS participants generally know the rules and thus probably feel a smaller inclination to explore them. Though this could possibly be a factor with influence, it does not explain why there is clustering in eye-movement data across participants. In fact, the decreased inter-participant variability among experienced observers suggests that a common strategy was utilized. It is quite likely that studying NKS and CA images biased the experienced observers to notice similar properties in CA images. Moreover, having prior knowledge about what kinds of images will be presented probably adds some level of top-down expectations.

One such top-down attentional bias among experienced observers is that all of their eye-movement data is not only clustered but also is closer to the top of the screen. Since the experienced participants know that the CA images are actually time-evolutions of 1-D CAs, there is a bias to start at the top and examine the earlier time steps. This attentional bias emerges due to prior knowledge and experience. Such attentional biases and other common strategies that emerge due to a similar knowledge base can account for the decrease in inter-participant variability within the experienced group. Moreover, the fact that the clustering did not take place in images containing an array of CA images suggests that the clustering effect was a consequence of previously developed top-down biases towards CAs (see Figure 9).

5. Conclusion

In conclusion, analysis of eye-tracking data from cellular automaton (CA) images provided many interesting results. For simple and random images it was noticed that a lack of an obvious hierarchy of salience leads to greater levels of inter-participant variability and subsequently implies greater top-down or extraneous influence.
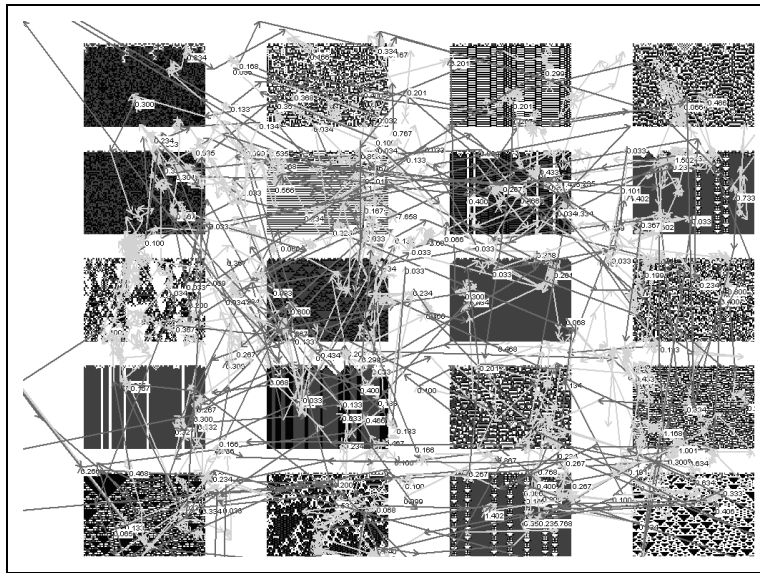
(a)
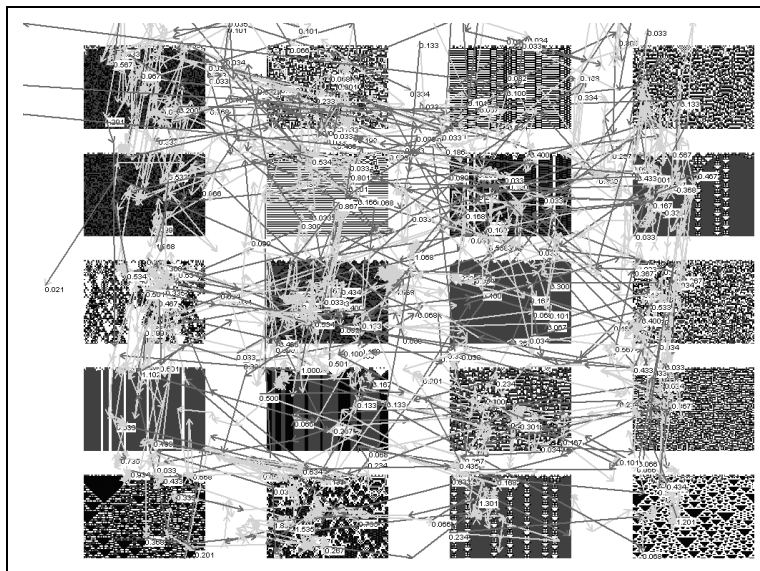
(b)

Figure 9. The clustering effect was not noticed for any of the four arrays of CA images. The fact that clustering was specifically limited to individual CA images suggests that the source of the clustering is due to top-down factors resulting from previously developed attentional biases towards CA images rather than due to other extraneous confounds such as their common academic backgrounds.

For random images with pockets of regularity, it was noticed that those pockets enjoyed fixation from the majority of participants. This finding raised questions about the exact relationship between higher-level structural features such as randomness or regularity, and questions about the kinds of visual features that can be used for discerning salience. The answers to these questions still remain unclear and thereby provide an interesting avenue for further work.

The eye-movement data collected for nested images were quite surprising, and escaped explanation within current oculomotor models. The hypotheses presented in this paper as explanations of the findings greatly relied upon the idea of the economical nature of the visual system. The extent to which the "economical" nature plays a role in determining patterns of fixation is quite unclear. Moreover, the kinds of features that such an economical system would take into consideration are currently based on speculation. Significant future exploration is required to explain the eye-movement data from nested images.

Finally, differences in oculomotor strategies between the naïve group and the experienced group not only confirmed previous results regarding the impact of familiarity on visual salience, but also reaffirmed the ease and subtlety with which top-down factors become relevant and also the sharp influence on oculomotor strategy that they can have. These data also suggest the difficulty experienced in creating situations where one can be entirely sure that the top-down and bottom-up mechanisms were completely isolated. Whether this is even possible, still remains open to debate.

In the end, it is clear that most of the above mentioned results do not match predictions from known models of human oculomotor strategy. Hence, the main assertion of this study is that further eye-tracking experiments using CA images could greatly benefit our current understanding of the human visual perceptual system.

5.1 Limitations of this study

There were a few concerns about the design of this study. First, the two groups of participants were not balanced for sex or handedness. However, there is no evidence in the literature suggesting that sex or handedness affects oculomotor strategy.

Second, there was concern that, without any specific task to perform, the participants might not stay interested. This concern was addressed by selecting images that would provide features to capture visual attention. Moreover, even though the experienced participants had seen images of similar type with similar characteristics, there was an explicit attempt to ensure that a large number of the CA images were novel even to the experienced group.

Third, the fact that all images were presented in the same order to all participants raised some concerns about order effects. It is possible that images presented later were treated with greater impatience and lack of interest. However, since the different types of CA images (simple, nested, random, and localized structures) were presented in random order, any potential order effects should have minimal impact on the observations presented here.

Finally, though this study was conducted specifically with the intention of discovering trends that would motivate future research studies, the lack of statistical data makes it difficult to make strong assertions.

5.2 Areas of future interest

As stated before, an examination of the exact relationship between higher-level stimuli features such as regularity or randomness and the mechanism for determining saliency is required.

Such investigations would allow speculations on how stimulus features such as randomness and regularity are represented, recognized, and chosen to be relevant features. Moreover, given the visual saliency model's assumption of a single saliency map, such studies would shed more light on which visual criteria can be used to determine saliency in a given situation.

It would also be useful to examine the extent and manner in which the nested structure of an image influences the resulting patterns of fixation. Further investigations should emphasize providing a defensible explanation for the occurrence of fixation points in blank regions. One possible way to test the above-proposed explanation based on conspicuity area is to determine if the saccades are systematically longer after a fixation in a blank region. If so, that would support the argument that fixation in blank regions allowed for the parallel processing of a larger amount of information.

It seems that any truly plausible explanation for this phenomenon would also have to give rise to a better understanding of how fixation patterns emerge. More specifically, it would have to illuminate the degree to which the economical nature of visual perception affects the patterns of fixations and the choice of fixation locations.

If there is no evidence for an economical visual system that guides fixation patterns, then different motivating factors need to be put forth so as to provide a plausible explanation for the results described here.

Acknowledgments

References

[1] S. Wolfram, *A New Kind of Science* (Wolfram Media, Inc., Champaign, IL, 2002).

[2] L. Itti and C. Koch, "A Saliency-based Search Mechanism for Overt and Covert Shifts of Visual Attention," *Vision Research*, **40** (2000) 1489–1506.

[3] W. S. Geisler and K. L. Chou, "Separation of Low-level and High-level Factors in Complex Tasks: Visual Search," *Psychological Review*, **102**(2) (1995) 356–378.

[4] M. M. Chun and R. Marois, "The Dark Side of Visual Attention," *Current Opinion in Neurobiology*, **12** (2002) 184–189.

[5] D. Parkhurst, K. Law, and E. Niebur, "Modelling the Role of Salience in the Allocation of Overt Visual Attention," *Vision Research*, **42** (2002) 107–123.

[6] A. A. Kustov and D. L. Robinson, "Shared Neural Control of Attentional Shifts and Eye Movements," *Nature*, **384** (1996) 74–77.

[7] K. A. Turano, D. R. Geruschat, and F. H. Baker, "Oculomotor Strategies for the Direction of Gaze Tested with a Real-world Activity," *Vision Research*, **43** (2003) 333–346.

[8] G. Z. Yang, L. Dempere-Marco, X. P. Hu, and A. Rowe, "Visual Search: Psychophysical Models and Practical Applications," *Image and Vision Computing*, **20** (2002) 291–305.

[9] B. C. Motter and E. J. Belky, "The Guidance of Eye Movements During Active Visual Search," *Vision Research*, **38** (1998) 1805–1815.

[10] F. H. Hamker, "A Dynamic Model of How Feature Cues Guide Spatial Attention," *Vision Research*, **44** (2004) 501–521.

[11] M. F. Land and M. Hayhoe, "In What Ways Do Eye Movements Contribute to Everyday Activities?" *Vision Research*, **41** (2001) 3559–3565.

[12] R. M. McPeek and E. L. Keller, "Short-term Priming, Concurrent Processing, and Saccade Curvature During a Target Selection Task in the Monkey," *Vision Research*, **41** (2001) 785–800.

[13] S. Suzuki and P. Cavanagh, "Facial Organisation Blocks Access to Low-level Features: An Object Inferiority Effect," *Journal of Experimental Psychology: Human Perception and Performance*, **21**(4) (1995) 901–913.

[14] B. C. Motter and E. J. Belky, "The Zone of Focal Attention During Active Visual Search," *Vision Research*, **38**(7) (1998) 1007–1022.