

# Peripheral vision and crowding in mental maze-solving

**Yelda Semizer**

Department of Humanities and Social Sciences, New Jersey Institute of Technology, Newark, NJ, USA



**Dian Yu**

Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA



**Ruth Rosenholtz**

Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA  
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA



**Solving a maze effectively relies on both perception and cognition. Studying maze-solving behavior contributes to our knowledge about these important processes. Through psychophysical experiments and modeling simulations, we examine the role of peripheral vision, specifically visual crowding in the periphery, in mental maze-solving. Experiment 1 measured gaze patterns while varying maze complexity, revealing a direct relationship between visual complexity and maze-solving efficiency. Simulations of the maze-solving task using a peripheral vision model confirmed the observed crowding effects while making an intriguing prediction that saccades provide a conservative measure of how far ahead observers can perceive the path. Experiment 2 confirms that observers can judge whether a point lies on the path at considerably greater distances than their average saccade. Taken together, our findings demonstrate that peripheral vision plays a key role in mental maze-solving.**

## Introduction

Solving a maze entails navigating through a network of passages with the objective of finding the path from an entry point to an end point. To solve a maze effectively, one needs to perceive the path ahead and aspects of the maze layout, note dead-ends, explore paths systematically, and backtrack if needed. As such, maze-solving is a multilayered process involving perception, action, and cognition (Zhao, Marquez, Hemmer, & Kowler, 2013). This article focuses on perceptual aspects.

Solving an even moderately complex maze cannot be done at a glance. Nor does maze-solving involve tracing every inch of the path by eye. Rather, maze-solving involves a discrete set of fixations on or near the path (Crowe, Averbeck, Chafee, Anderson, & Georgopoulos, 2000). How might the observer saccade in order to mentally solve a maze?

Previous work on saccade planning and execution suggests that saccades are used to improve the visibility of task-relevant information in tasks as diverse as visual search (e.g., Najemnik & Geisler, 2005; Najemnik & Geisler, 2009), reading (e.g., Legge, Klitz, & Tjan, 1997), and shape recognition (e.g., Renninger, Verghese, & Coughlan, 2007). For example, Najemnik and Geisler (2005) asked observers to search for a sine-wave grating embedded in  $1/f$  noise and found that the number of saccades to locate a target and the location of fixations were predicted by a Bayesian ideal observer model, limited by the fall-off in the target detectability as a function of eccentricity, which directs gaze to locations with highest target probabilities. Their later work (Najemnik & Geisler, 2009) showed that gaze patterns were also predicted by an ideal observer using a simpler but biologically more plausible strategy (i.e., entropy limit minimization), which directs gaze to locations, minimizing the expected uncertainty.

We might extend this sort of logic to saccade planning in maze-solving tasks as follows: Starting at the maze entrance, an observer fixates a point along the path while likely extending their visual processing along the path as far as possible. Based on the visual information available at the current fixation, the observer iteratively chooses a new fixation point to gain additional information until they perceive the

Citation: Semizer, Y., Yu, D., & Rosenholtz, R. (2024). Peripheral vision and crowding in mental maze-solving. *Journal of Vision*, 24(4):22, 1–13, <https://doi.org/10.1167/jov.24.4.22>.



exit. In this scheme, the degree to which the observer perceives the path ahead depends on peripheral vision, the visual processing beyond our current point of gaze. Peripheral vision can encounter uncertainty regarding the direction of the path either due to branching paths or because perceptual factors such as visual clutter significantly diminish our ability to discern the upcoming path. In such cases, a well-functioning visual system would shift the point of gaze to acquire additional information. Making a long saccade in the face of uncertainty risks deviating from the path. Consequently, when one's ability to perceive the path ahead is limited, one might prefer to execute shorter saccades. However, overly conservative choice of saccade length reduces maze-solving efficiency.

Visual crowding provides a primary limiting factor in the use of peripheral vision (Rosenholtz, 2016). Crowding refers to the harmful effect of clutter on one's ability to perceive information in the periphery (Bouma, 1970; Levi, 2008; Pelli et al., 2007). In fact, previous work on maze-solving and related tasks has found evidence for effects consistent with visual crowding. Ullman and colleagues (Jolicoeur, Ullman, & Mackay, 1991; Jolicoeur & Ingleton, 1991; Ullman, 1996) studied visual cognition tasks equivalent to maze-solving, such as tasks in which observers had to decide whether two points lay along a single line or distinct lines. Observers took longer to respond when the distance between the points was greater (Jolicoeur, Ullman, & Mackay, 1986), when the lines were in close proximity, or when the lines exhibited greater curvature (Jolicoeur et al., 1991), consistent with use of both peripheral vision in general and visual crowding in particular. Intuitively, curvature would increase crowding because curved stimuli are more complex and include more orientations. If crowding is due to some sort of feature averaging or pooling, then in the stimuli of the curve-following tasks of Jolicoeur et al. (1991); Jolicoeur and Ingleton (1991); Roelfsema (2006), and Ullman (1996), the average orientation in the stimuli composed of straight lines is the same as the stimulus orientation, so that maze will be well represented by the average, whereas in the stimuli composed of curved lines, the average orientation does not well represent the stimulus. Similarly, in visual crowding, identification performance declines when stimuli, such as letters, in the periphery are flanked by other letters that are in close proximity or in critical spacing (Bouma, 1970). Moreover, crowding can result in mislocalization errors (Korte, 1923), and the level of crowding is influenced not just by the distance to flanking stimuli but also the stimulus characteristics (Pelli, Burns, Farell, & Moore-Page, 2006; Pelli & Tillman, 2008). Indeed, Jolicoeur et al. (1991) attributed their results to crowding (referring to it as “lateral masking,” a term previously used somewhat interchangeably with “crowding”). However, Jolicoeur et al. (1991)

focused on one aspect of crowding, its dependence on spacing between stimulus items such as neighboring paths.

Similarly, a study by Crowe et al. (2000) on mental maze-solving (i.e., solving a maze mentally without marking it physically) showed that both the length of the path (i.e., the distance to be traversed between the maze entrance and exit) and the frequency of turns (i.e., 90° turns along the path) contributed to maze-solving performance and demonstrated that the duration of the current fixation depends upon both the length of the next saccade and the frequency of turns between the two fixations. That is, observers actively use the information that is available in their periphery while moving their eyes, once again implying a role of peripheral vision in mental maze-solving.

In a recent study, Yu, Wan, Balas, and Rosenholtz (2019) investigated the effects of maze appearance on maze-solving performance. By measuring or systematically manipulating the path and wall thickness as well as path style in simple mazes, authors showed that maze-solving time increases with thicker paths and wavy walls. These findings suggest that perceptual aspects of a maze influence maze-solving performance in a way consistent with visual crowding (Rosenholtz, 2016) and figure/ground segmentation. The current study builds upon the previous findings by examining eye movements using behavioral experiments and modeling.

A perceptually complex and visually crowded maze should increase the ambiguity associated with the path ahead, forcing the visual system to execute smaller saccades to resolve this ambiguity as they solve the maze. This would mean that for more crowded mazes, observers would make a larger number of fixations for a given path length. In Experiment 1, we manipulated the visual complexity of mazes while controlling other factors (e.g., path length, number of turns, etc.) and measured the fixations observers made to solve the mazes. We then modeled the maze-solving task using a model of peripheral vision, the Texture Tiling Model (TTM; Balas, Nakano, & Rosenholtz, 2009; Ehinger & Rosenholtz, 2016; Keshvari & Rosenholtz, 2016; Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Zhang, Huang, Yigit-Elliott, & Rosenholtz, 2015), to test whether the model can explain the observed fixation data. We show that the model consistently predicts a smaller number of fixations than executed by the observers in Experiment 1. This suggests an intriguing prediction: Observers might conservatively pick fixation locations and might actually perceive farther ahead than suggested by their saccades. To empirically test this prediction, Experiment 2 examined maze perception beyond typical saccades.

## Experiment 1: Does visual complexity influence eye movements while mental maze solving?

### Methods

#### Observers

A total of nine observers participated in the experiment. Two were authors of this article (S1 and S2); the rest were naive to the purpose of the experiment and were compensated for their participation. All observers had normal or corrected-to-normal visual acuity.

#### Apparatus

Stimuli were presented on a 55-in. LG OLED TV at 60 Hz with a resolution of  $1,920 \times 1,080$  pixels. The viewing distance of the observers was 82 cm and the display window subtended  $70^\circ \times 40^\circ$  of visual angle. MATLAB software (MathWorks, Natick, MA, USA) and the Psychophysics Toolbox extensions (Brainard, 1997) were used to present the stimuli. An Eyelink 1000 infrared eye tracker (SR Research, Kanata, Ontario, Canada) was used to monitor and record eye movements of the observers monocularly at 1,000 Hz. The head position of the observers was stabilized using a forehead- and chin-rest. Fixations were classified using a parsing algorithm (Geisler, Perry, & Najemnik, 2006) implemented in MATLAB (MathWorks).

#### Stimuli

Stimuli were composed of 24 images of two-dimensional mazes (see Figure 1). Each maze was composed of a  $13 \times 13$  array of square cells and the side of each cell subtended approximately  $1.8^\circ$  of visual angle. Mazes varied in terms of the length of the path and the number of turns along the path. The average path length was 41 square cells (range: 21–53) while the average number of turns was 25 (range: 14–38).

Original mazes were generated using an online maze generator ([www.mazegenerator.net](http://www.mazegenerator.net)) and were modified using Adobe Photoshop 20 (Adobe, San Jose, CA, USA) software as follows: (a) Original maze images had only one exit, which was always located at the top of the maze at the center. To motivate observers to mentally solve the maze rather than just passively view them, two more exits were added to each maze, and the path was edited so that only one of these exits was the correct solution. The location of the correct exit (left, middle, or right) was equally represented across mazes. (b) Original maze images had multiple

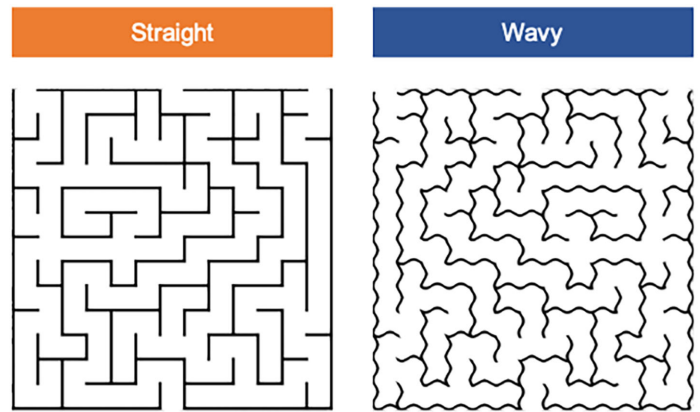


Figure 1. An example of a maze with straight (on the left) and wavy (on the right) walls. Original mazes were generated using an online maze generator ([www.mazegenerator.net](http://www.mazegenerator.net)) and were modified so that all maze images had a single entrance at the bottom at the center. Three possible exits were always located at the top, only one of which was the correct solution. Note that the wavy maze was flipped horizontally.

paths with dead-ends. Having such paths can make the maze-solving task more cognitively challenging (e.g., increased memory load due to a need to remember previously visited paths while backtracking). Since the primary focus of this study was to examine perceptual factors rather than cognitive factors, each maze was modified to have a single enclosed path. (c) Original maze images had straight walls. To manipulate the complexity of mazes, the walls of each maze were modified to be wavy (see Figure 1). Using the same maze topology in both conditions ensured that the two sets of mazes were otherwise equivalent (e.g., in terms of path length and number of turns). Maze images with wavy walls were flipped horizontally to minimize any advantage the observer might get from solving a maze with the same topology a second time. Observers reported that they did not notice that same mazes were used in two conditions.

#### Procedure

At the beginning of the experiment, observers received instructions explaining the task and the maze design (e.g., mazes had a single enclosed path with a single entrance at the bottom and three possible exits at the top). Observers were instructed to start solving the maze at the entrance and to do so as quickly as possible, relying only on their vision to navigate, without using any additional means of path tracing. At the start of each trial, observers fixated a green dot (see Figure 2). Observers pressed the start key to initiate the trial. Fixation tolerance was set so that the trial started only if observers were fixating within  $0.5^\circ$  of the green dot. Then, a randomly chosen maze image appeared at the

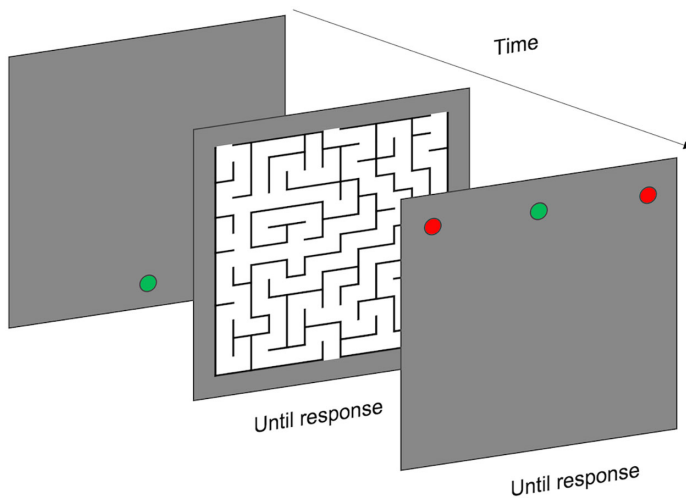


Figure 2. Maze-solving task sequence. The green dot on the first frame indicates the entrance location. Observers pressed a key to initiate the trial and display the maze while fixating at the green dot. They mentally solved the maze by moving their eyes and pressed a key to discard the maze. Then, three red dots were displayed to indicate possible exit locations. Observers looked at the dot representing the correct exit location and pressed another key to log their responses. The red dot nearest the current fixation was highlighted in real time to facilitate selection of the chosen location. In this example, the observer is looking at the center dot.

center of the screen, with its entrance at the location of the previously fixated green dot. Observers were given unlimited time to solve the maze. After observers pressed a key to indicate they solved the maze, the image disappeared, and three red dots appeared on the screen to mark the locations of possible exits. Observers were asked to look at the point corresponding to the chosen exit location and press a key to log their response. The red dot nearest the current fixation was highlighted in real time to facilitate selection of the chosen location.

We recorded the fixations made while solving the maze as the primary measure of performance. The first fixation, which was always at the entrance of the maze, was excluded from the analysis.

Trials were blocked by the maze condition (straight or wavy). The order of blocks was counterbalanced across observers. Each block consisted of 3 practice trials and 24 experimental trials. Data from the practice trials were not included in the analysis. At the start of each block, observers were required to complete a 9-point calibration routine. If necessary, the calibration routine could be repeated during the block. If the observers blinked while solving the maze, the trial was aborted, the data were discarded, and the observers were informed. The image from the aborted trial was repeated later in the experiment. The experiment took approximately half an hour to complete.

## Results

### Accuracy

Observers were highly accurate in choosing the correct exit after solving the mazes, with accuracy levels above 99%. Although this result is expected because the maze-solving task was self-paced, it confirms that observers indeed solved the mazes. There was only one trial where an observer made an inaccurate response. Eye movement analysis suggested that the observer made only two fixations at the start of the maze path and terminated the trial prematurely. This data point has been excluded from further analyses.

### Path efficiency

Our main hypothesis was that observers will solve wavy mazes less efficiently than straight mazes due to a decrease in their ability to see in the periphery. To measure the efficiency of path traversal, we computed path efficiency by dividing the path length of each maze by the number of fixations made by an observer while solving the maze. Figure 3 shows the average path efficiency for each observer across two conditions. A one-way repeated-measures analysis of variance (ANOVA) revealed a significant effect of maze condition,  $F(1, 8) = 17.0$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.68$ . Observers were significantly less efficient while solving wavy mazes ( $M = 2.60$ ,  $SE = 0.12$ ) compared to straight mazes ( $M = 2.95$ ,  $SE = 0.18$ ). This suggests that observers took shorter “steps” along the path to solve the wavy mazes, given that the path lengths were the same across two conditions.

### Saccade amplitudes

The above analysis focuses on the efficiency of path traversal. Given that the path lengths were the same across conditions, the analysis suggests that observers took shorter “steps” along the path while solving wavy mazes compared to straight mazes. We tested whether a similar relationship also exists in the saccade amplitudes (i.e., how far the observer moves their eyes in degrees of visual angle [dva]). It is possible that observers not just move their eyes less distance along the path, in the wavy mazes, but also make shorter saccades. For example, if there is a U-shape portion of the path, an observer might make a single eye movement from one end of the U to the other, resulting in a saccade length that is considerably shorter than the path length between the two end points.

We measured the saccade amplitudes by computing the Euclidean distance between two consecutive fixation locations. Then, we tested whether the average saccade amplitude differs between two conditions using a one-way repeated-measures ANOVA. Results

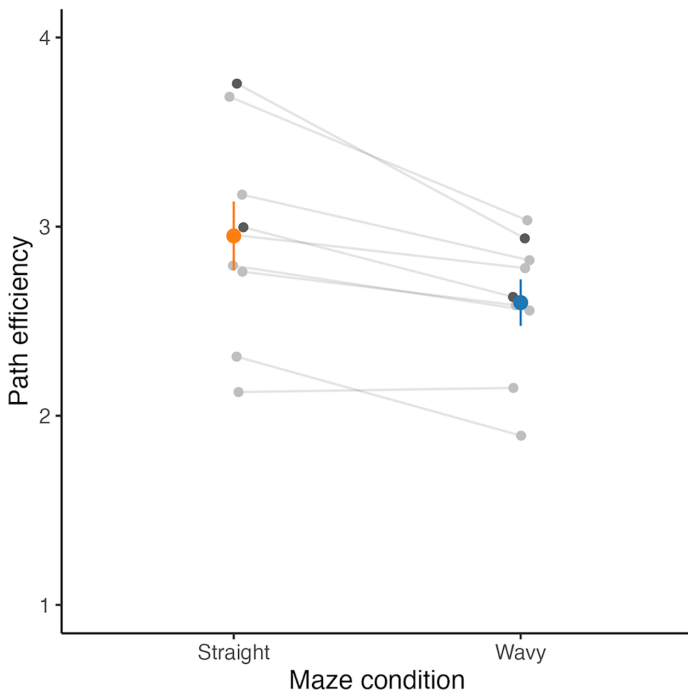


Figure 3. Average path efficiency in straight and wavy conditions. Path efficiency was computed by dividing the path length of each maze by the number of fixations made by an observer while solving the maze. Light gray dots connected with lines indicate the average path efficiency for individual observers; orange and blue dots indicate the average performance in the straight and wavy conditions, respectively; and error bars indicate the standard error. Dark gray dots indicate the performance for S1 and S2, who were also authors of this article. Observers solved wavy mazes less efficiently compared to straight mazes on average.

showed a significant effect of maze condition,  $F(1, 8) = 6.60, p = 0.03, \eta_p^2 = 0.45$ , suggesting that observers made shorter saccades while solving wavy mazes ( $M = 2.89, SE = 0.17$ ) compared to straight mazes ( $M = 3.05, SE = 0.20$ ). This suggests that observers also traveled “shorter” distances while choosing their fixation locations in wavy mazes compared to straight mazes.

### Duration of fixations

It is possible that an observer dwells longer at a given fixation location in order to accumulate more information and then makes a more informed and longer saccade. If so, more fixations when solving the wavy mazes might not indicate greater difficulty perceiving the path ahead in a more crowded maze. One should ask, then, whether observers displayed longer fixation durations for the straight mazes, where they make longer saccades, compared to wavy mazes.

We computed the average fixation duration in each trial per observer. We then analyzed these across

two conditions using a one-way repeated-measures ANOVA, which revealed that fixation durations did not differ significantly ( $p > 0.05$ ) between the wavy ( $M = 0.20, SE = 0.27$ ) and straight ( $M = 0.20, SE = 0.29$ ) mazes. These results suggest that the observed difference in saccade amplitudes cannot be explained by differences in fixation duration.

### Path length

Based on prior work (Crowe et al., 2000), one would expect the number of fixations to increase as a function of increasing path length. Figure 4 (left) shows the median number of fixations observers made while solving each maze as a function of path length measured as the number of square cells one has to travel to solve the maze. In both conditions, the number of fixations increases as a function of path length,  $r(22) = 0.93, p < 0.001$  and  $r(22) = 0.94, p < 0.001$ , in straight and wavy conditions, respectively, but wavy mazes still result in a larger number of fixations than straight mazes.

### Number of turns

An additional interesting question is whether observers were selective in choosing their fixation locations along the path. For example, one naive way of thinking of this problem is to assume that observers will look at the next turning point while following the path. Indeed, there is prior evidence that suggests increasing number of fixations with increasing number of turns (Crowe et al., 2000). Figure 4 (right) shows human performance as a function of the number of turns in each maze. Although the number of fixations correlated with the number of turns within the path,  $r(22) = 0.92, p < 0.001$  and  $r(22) = 0.85, p < 0.001$ , in straight and wavy conditions, respectively, observers still made more fixations while solving mazes with wavy walls compared to straight walls, which suggests that the performance cannot be explained just by the number of turns. While observers make frequent saccades, they do not necessarily fixate on every turn. Based on fixation counts, they are consistently able to see farther ahead than the upcoming turn and saccade accordingly. Taken together, the differences in the appearance of the maze walls contributes to performance above and beyond the factors of path length and number of turns.

**Does a model of peripheral vision make sense of these eye movements?**

To test whether a model of peripheral vision, the TTM (Balas, et al., 2009; Rosenholtz, Huang, &

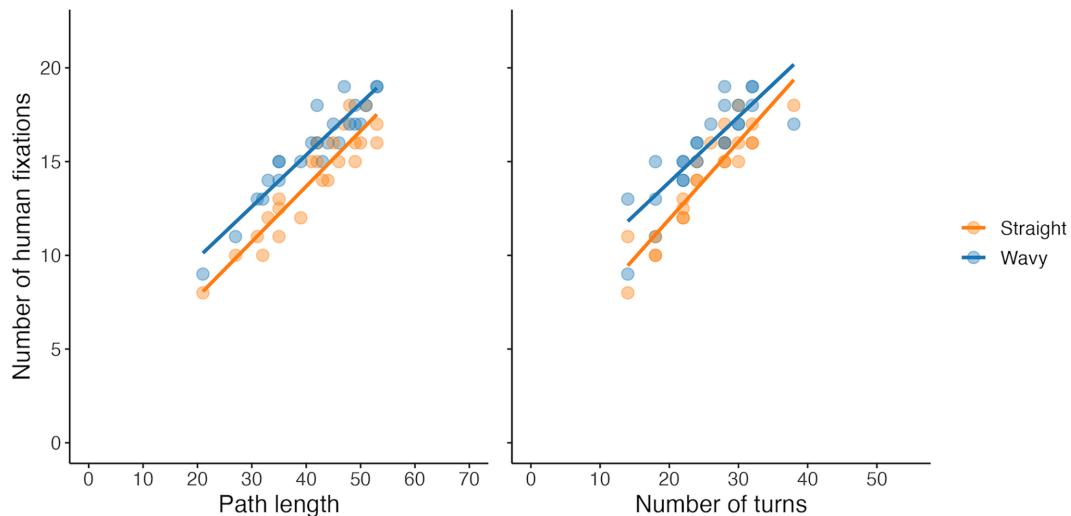


Figure 4. Left: Median number of human fixations as a function of path length for each maze in two conditions. Right: Median number of human fixations as a function of number of turns for each maze in two conditions. Orange dots represent the straight condition and the blue dots represent the wavy condition. Lines show least squares fits. Number of fixations increases as the path length and the number of turns increase in both conditions, but wavy mazes still result in a larger number of fixations than straight mazes on average.

Ehinger, 2012; Rosenholtz, Huang, Raj et al., 2012), can be used to make sense of human fixations, we ran model simulations using the maze images used in Experiment 1. Specifically, we were interested in measuring how far ahead the model sees a clear path while fixating a point along the path, starting from the beginning of the path (i.e., the entrance of the maze). Assuming that the model is going to “saccade” to the location where the path becomes unclear, model simulations were computed.

For each maze in our stimulus set, we generated multiple “mongrels” (Balas et al., 2009) using a fixation point (i.e., first fixation point was always the maze entrance). “Mongrels” are images that visualize the information preserved and lost in peripheral vision, according to the model. The visualizations generated from a single maze and fixation point can show some variability. Information that is clearly available in multiple mongrels—such as whether the maze path continues in a certain direction—the model predicts is readily available to peripheral vision. Figure 5 shows an example maze with three mongrel images. After inspecting each mongrel image, we chose the next fixation point by tracing the path starting from the current fixation point and checking how far a path was clearly visible. When there was a point where multiple routes were possible or the path appeared blocked, we marked this point where the path became unclear, and then we chose to have a new fixation. Using this new point as the next fixation point, we generated a new set of mongrel images of the original maze image and repeated the same procedure until the model traveled

the entire path and the last model fixation led to a clear view of the exit. The number of simulated fixations for each maze was counted as the predicted number of fixations required to solve that maze. This whole process took several hours to complete as the model simulations took quite a long time to generate the next set of simulated images given the previous fixations; this serial process of iterating between slow model syntheses and human input required using the authors to select the next model fixation rather than having naive subjects make this judgment in an experimental setup.

## Results

The model predicted significantly more fixations in the wavy maze condition ( $M = 7.04$ ,  $SE = 0.41$ ) compared to the straight maze condition ( $M = 5.29$ ,  $SE = 0.30$ ),  $t(23) = 9.08$ ,  $p < 0.001$ ,  $d = 1.85$ . Figure 6 shows the number of model fixations versus human fixations for each maze in the straight and wavy conditions. The model underpredicted the number of fixations required to solve mazes in both conditions. We will discuss possible reasons for this later in the Discussion section. However, the number of model fixations correlates with the number of human fixations both in the wavy condition,  $r(46) = 0.69$ ,  $p < 0.001$ , and in the straight condition,  $r(46) = 0.76$ ,  $p < 0.001$ , suggesting that perceptually difficult mazes for humans were also difficult for the model.

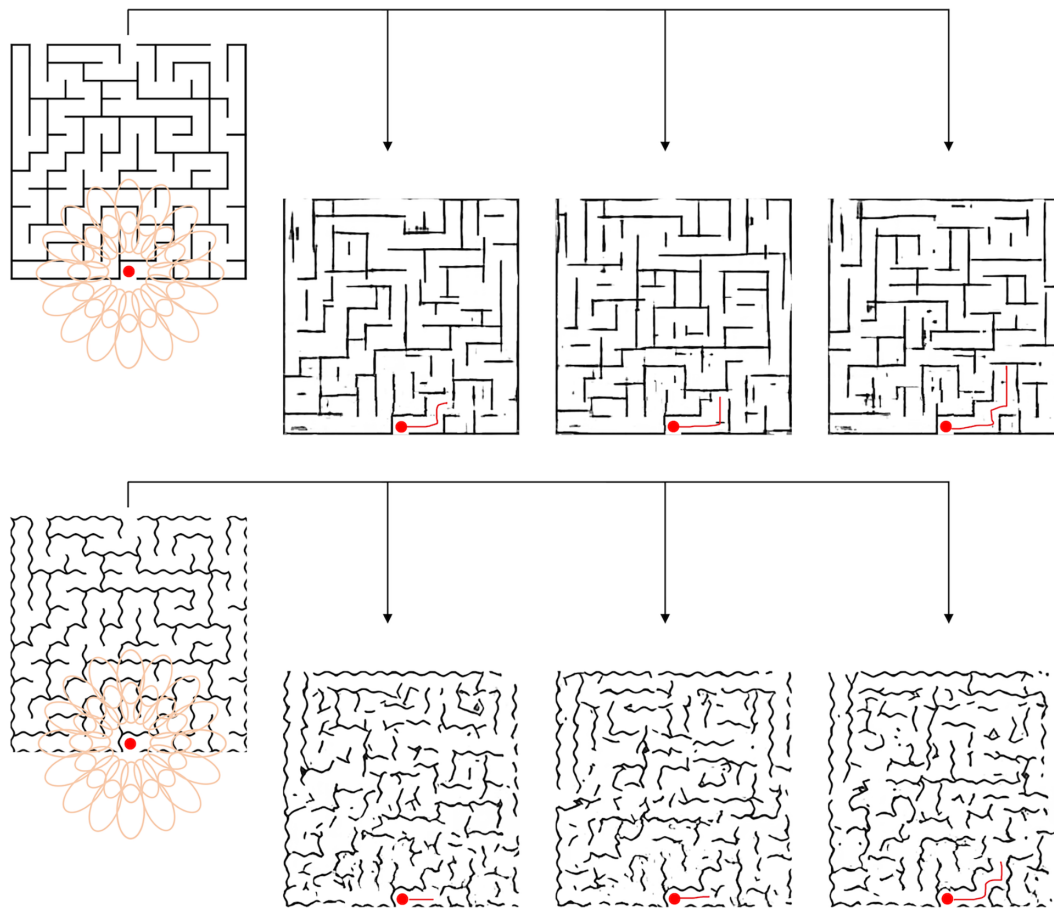


Figure 5. Mongrel images generated from a single maze image with straight (top row) and wavy (bottom row) walls. Fixation locations are represented with a red point and the possible clear paths are represented with red lines (the red points/lines are used for demonstration purposes and were not visible to the model).

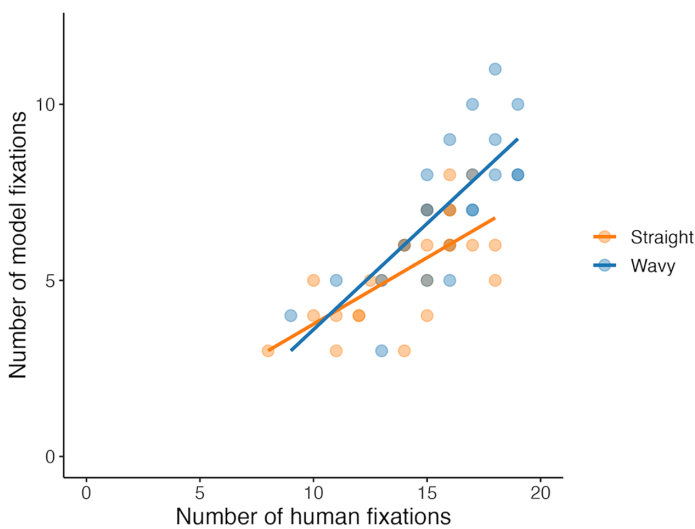


Figure 6. Human versus model fixations in the straight (orange) and wavy (blue) conditions. Each data point represents the average data for a single maze. Lines represent the least squares fits.

### Experiment 2: Can observers perceive paths farther along than their fixations?

Results of the model simulations underpredicted the number of fixations required to solve mazes. This finding could be due to several different factors. First, it is possible that TTM makes poor predictions for this type of stimuli and task. Earlier implementations of TTM successfully predicted crowding performance for recognition tasks using simple stimuli such as letters and symbols (Balas et al., 2009; Keshvari & Rosenholtz, 2016; Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, et al., 2012), or getting the gist of a real-world scene (Ehinger & Rosenholtz, 2016). Possibly the summary statistics computed by TTM require updating to deal with maze-like stimuli. Second, although the TTM accounts for the information fall-off in the periphery, other factors could influence task performance (e.g., motivation, fatigue, etc.). Third, it is

possible that the visual system might be conservative; in other words, observers can see farther along the path than indicated by their fixation locations. On the one hand, in sparse displays, when the observer makes a saccade toward a single isolated target, saccades tend to predominantly land on the target (Kowler & Blaser, 1995). On the other hand, for denser displays, such as  $1/f$  noise (e.g., Najemnik & Geisler, 2005) or natural scenes (e.g., Henderson & Hollingworth, 1998), saccades tend to be limited in their length. What we observed in our data aligns closely with the saccade lengths in the latter.

Note that due to the importance of path length in maze solving, our measure of path efficiency quantifies saccade length along the path of the maze, not in dva “as the crow flies.” Observers saccade on average 2.6 units along the maze path, whereas the model predicted observers could make an average saccade of 7.1 units along the path.

In Experiment 2, we ask whether observers can make correct judgments about the path at those more distant model fixations. If so, that implies that observed saccade lengths underrepresent ability to perceive the maze at a glance.

## Methods

### Observers

A total of nine observers participated in the experiment. Observers were naive to the purpose of the experiment and were compensated for their participation. All observers had normal or corrected-to-normal visual acuity.

### Apparatus

Stimuli were presented on a 27-in. BenQ monitor at 60 Hz with a resolution of  $3,840 \times 2,160$  pixels. The viewing distance of the observers was 70 cm and the display window subtended  $26^\circ \times 46^\circ$  of visual angle. MATLAB software (MathWorks) and the Psychophysics Toolbox extensions (Brainard, 1997) were used to present the stimuli. An EyeLink 1000 infrared eye tracker (SR Research) was used to monitor and record observers’ eye movements monocularly at 1,000 Hz. Observers’ head position was stabilized using a forehead- and chin-rest.

### Stimuli

The same maze images used in Experiment 1 were also used in Experiment 2. Since the goal of the experiment was to test whether observers can perceive the peripheral path at locations where the model fixated, rather than to compare straight versus wavy mazes,

we only used mazes from the straight condition. A black disk was used to indicate the target location (see Figure 8). The size of target disk was  $1^\circ$  of visual angle.

The target and fixation locations were given by the model simulations and varied for each stimulus (except for the first fixation location that was always at the entrance of the maze). For each maze, the model fixation  $n$  was used as the on-path target location while the observer fixated the previous model fixation ( $n - 1$ ). For each on-path target location, we picked an off-path target location by randomly choosing from the eight surrounding cells, with the restriction that the chosen location was off the path.

### Procedure

At the start of the experiment, observers were provided with a set of instructions on the screen explaining the task, the maze design (e.g., there is only one enclosed path that starts at the bottom and moves toward to top where there is an exit), and other task procedures accompanied by Figure 7, which explained the on-path and off-path manipulation. We used a two-interval forced-choice task. At the start of each trial, observers fixated a green dot representing the fixation location (see Figure 8). Observers initiated the trial by pressing the start key. The trial started only if observers were fixating within  $0.5^\circ$  of the fixation location. After a random stimulus-onset asynchrony (SOA) ranging from 100 to 300 ms. The stimulus sequence consisted of two stimulus displays presented for 500 ms each, with a blank gray display in between presented for 800 ms. One of the two stimulus displays contained a maze image with an on-path target while the other contained the same maze image with an off-path target. The task was to decide which of two stimulus displays contained the maze with the on-path target while maintaining fixation. Observers were required to maintain fixation while the green fixation dot was

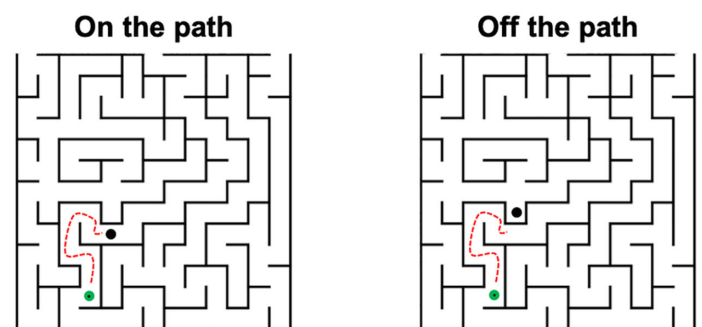


Figure 7. The image used during the instruction phase to explain the on-path and off-path manipulation. Red dashed line was used to indicate the path and was only shown during the instructions. In this example, the path length between the fixation dot and the target was eight cells.



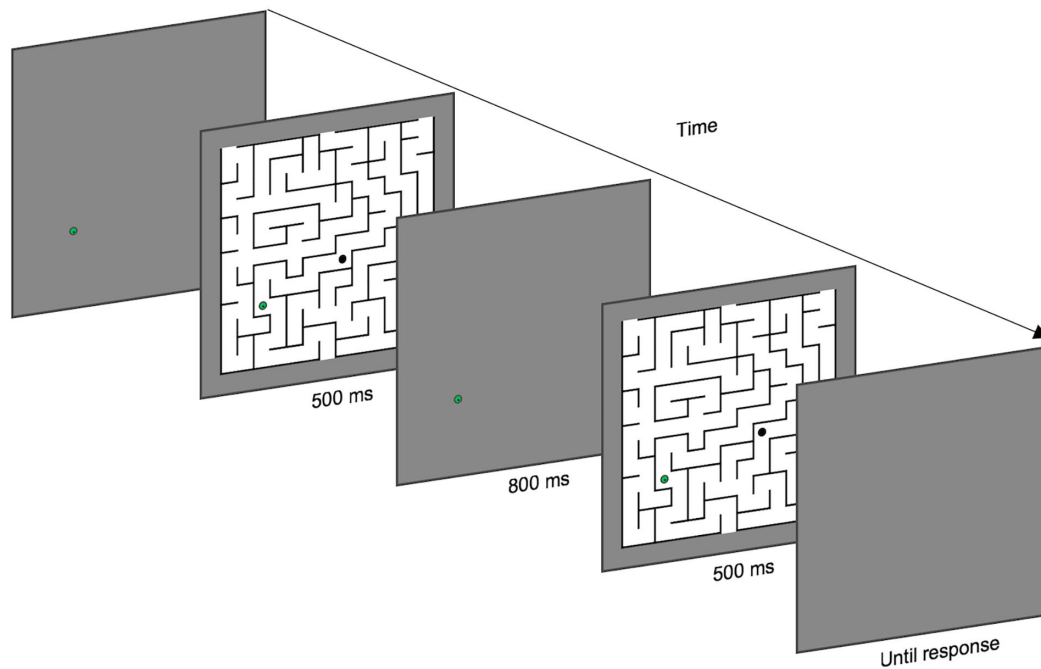


Figure 8. Detection task sequence. A green dot indicating the fixation location was presented on the screen. Observers pressed a key to initiate the trial while keeping their fixation. After a random SOA (100 to 300 ms), two stimulus displays presented for 500 ms each, with a blank gray display in between presented for 800 ms. Each maze image had an embedded target (black disk) either at an on-path or an off-path location. Observers responded with a key press to indicate which display had the on-path target.

on the screen (i.e., for the duration of two stimulus displays plus the gray screen in between). Observers were instructed to respond by key press as quickly and as accurately as possible. The response keys for first or second display judgments were counterbalanced across observers. The order of on-path and off-path stimulus was randomized for each trial across participants, as well as the order of trials. To minimize the effects of seeing the same maze stimuli multiple times, the maze image as well as the on-path and off-path stimuli were randomly flipped horizontally in each trial. The accuracy of the response was determined by comparing the selected interval to the target interval and was the primary measure of analysis.

Model simulations for the straight mazes resulted in a total of 103 fixations, combined across 24 mazes. For six of these fixations, it was not possible to choose an off-path location so these fixations were excluded, resulting in 97 testable model fixations. Trials were randomly ordered; repeated three times, resulting in three blocks and a total of 291 trials; and took approximately an hour to complete.

Observers completed 30 practice trials at the beginning of the experiment to familiarize themselves with the general task procedure. To ensure that observers were comfortable with maintaining fixations, the practice trials were repeated until all were completed without any break of fixations.

Data from the practice trials were excluded from the analysis.

At the start of each block, observers were required to complete a 9-point calibration routine. If necessary, the calibration routine could be repeated during the block. If an eye movement or a blink was detected during stimulus presentation, the trial was aborted immediately, the response was not collected, and the observers were informed. The stimulus from the aborted trial was repeated later in the experiment.

## Results

### Sensitivity

The main goal of [Experiment 2](#) was to test whether observers were above chance levels in classifying targets placed on locations at larger distances than they saccade to. We computed the average sensitivity of each observer in detecting targets using  $d'$  given by

$$d' = \Phi^{-1}(H) - \Phi^{-1}(F)$$

where  $H$  is the hit rate,  $F$  is the false alarm rate, and  $\Phi^{-1}$  is the inverse cumulative distribution function of the normal distribution. The performance at the chance level would be 50% accuracy (or  $d' = 0$ ). All observers performed above chance levels. We chose a higher

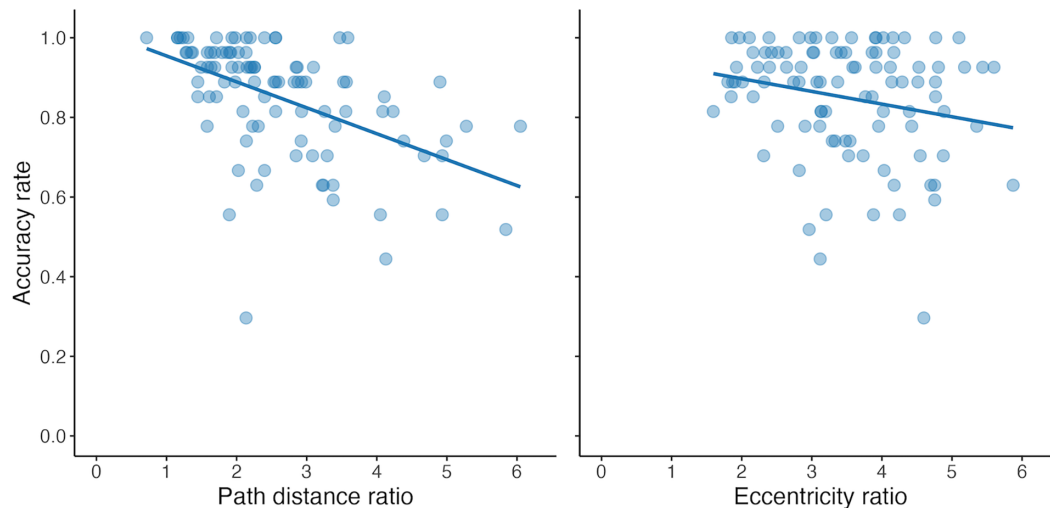


Figure 9. Left: Average accuracy as a function of path distance ratio (i.e., the ratio of the path distance of the target to the average saccade length for each maze). Right: Average accuracy as a function of eccentricity ratio (i.e., the ratio of eccentricity of the target to average saccade length for each maze) using on-path stimuli. Blue dots represent average accuracy for each target-fixation pair. Lines represent the least squares fits. Note that, a value of 4 on the x-axis implies that observers made a judgment about the path at an eccentricity four times greater than their typical saccade length for that maze.

performance level to be more conservative and tested the performance against  $d' = 1$  (69% accuracy) with a one-sample  $t$ -test, which revealed that the average  $d'$  ( $M = 2.09$ ,  $SE = 0.09$ ) was above 1,  $t(8) = 12.27$ ,  $p < 0.001$ ,  $d = 4.09$ . These findings suggest that observers were highly accurate in detecting targets placed at locations at greater distances than their typical saccade lengths.

### Target distance

We also analyzed how performance changes as a function of target distance, that is, the distance between the target location and the fixation location, using the on-path stimuli. We expected a decrease in average accuracy as a function of an increase in target distance. We quantified target distance in two ways: path distance and eccentricity. Path distance was computed by counting the number of cells along the path between the fixation point and the target point. Eccentricity was computed by quantifying the distance between the target point and the fixation point in dva.

Then, we compared the size of these target distances to typical saccade lengths from Experiment 1 by computing ratios as follows. Path distance ratio was computed by taking the ratio of path distance to the average saccade length in units of cell (given by the path efficiency measure). Eccentricity ratio was computed by taking the ratio of eccentricity to average saccade amplitude in dva. If the ratio is close to 1, it suggests that the target was positioned at a distance similar to the average length of a saccade. If the ratio is larger than 1, it suggested that the target was positioned

at a distance larger than the average length of a saccade.

Figure 9 shows performance as a function of path distance ratio and eccentricity ratio. Blue dots represent average accuracy for each target-fixation pair. A value of 4 implies that observers made a judgment about the path at an eccentricity four times greater than their typical saccade length for that maze. Average accuracy decreases as path distance ratio increases,  $r(95) = -0.51$ ,  $p < 0.001$ , and as the eccentricity ratio increases,  $r(95) = -0.22$ ,  $p = 0.03$ . The Fisher-transformed  $z$ -scores indicated a significant difference in the observed correlations,  $z = -2.41$ ,  $p < 0.05$ . These findings suggest that observers can classify maze locations as being on or off the solution path when those locations appear at a greater distance than typical saccades.

## Discussion

Mentally solving a maze requires following a path that connects an entrance to an exit without physically marking it. Anticipating the next eye movement requires getting a sense of the path beyond the current point of gaze, a process that involves peripheral vision. When the path ahead is cluttered, the visual system might have less confidence about the path direction farther ahead, leading the observer to execute shorter saccades. Motivated by these insights, the goal of the present study was to investigate the role of peripheral vision in a mental maze-solving task. We tested the impact of visual crowding on the efficacy of mental maze-solving,

using behavioral experiments and modeling simulations. Our results suggest that peripheral vision facilitates mental maze-solving. Our main findings are as follows:

First, mazes characterized by increased visual complexity, achieved by altering the appearance of maze walls, forced observers to make shorter saccades along the path, leading to more fixations to solve the maze, for a given path. These results agree with intuitions about crowding, in which additional complexity or clutter lead to poorer peripheral performance, which in maze-solving would lead to greater uncertainty as to the upcoming path.

Second, the crowding effects observed in [Experiment 1](#) were confirmed by a model of peripheral vision. The simulation of the mental maze-solving task using the TTM ([Balas et al., 2009](#); [Ehinger & Rosenholtz, 2016](#); [Keshvari & Rosenholtz, 2016](#); [Rosenholtz, Huang, & Ehinger, 2012](#); [Rosenholtz, Huang, Raj, et al., 2012](#); [Zhang et al., 2015](#)) showed that the model made more fixations for the wavy mazes than for the straight. However, the model made considerably fewer fixations compared to humans.

The variance in the fixation counts between human and model performance could be due in part to the lack of prior testing of TTM on similar stimuli. Prior applications of the TTM predominantly used letters or symbols accompanied by simple recognition tasks ([Balas et al., 2009](#); [Keshvari & Rosenholtz, 2016](#); [Rosenholtz, Huang, & Ehinger, 2012](#); [Rosenholtz, Huang, Raj, et al., 2012](#)), or natural scenes and scene gist tasks ([Ehinger & Rosenholtz, 2016](#)). However, aspects of the saccade planning system may provide a more parsimonious explanation. For example, human searchers tend to make shorter saccades than model searchers in certain conditions ([Najemnik & Geisler, 2009](#)). Making an error and deviating from the path would delay maze-solving, whereas saccades take little time. This may lead the visual system to adopt a conservative approach to selecting fixation locations.

[Experiment 2](#) provided evidence for conservative fixation selection. Human observers were indeed capable of perceiving the difference between on- and off-path targets positioned at the greater distance suggested by the model fixations, even though those fixations were on average 4.5 units farther along the path than the observer fixations (a 170% greater distance along the path, on average).

Further analysis of data from [Experiment 2](#) demonstrated that ability to judge whether a peripheral target lay on the path decreased as a function of the path length, measured in terms of number of square cells between the target and fixation locations. While performance also decreased as a function of eccentricity of the target, the effect was considerably weaker. This result is consistent with previous work showing that time to solve a curve-following task ([Jolicoeur et al.,](#)

[1986](#)) or a maze task ([Crowe et al., 2000](#)) depends on path length.

It might be true that observers in our study preferred to make additional eye movements while solving mazes rather than relying on difficult peripheral judgments. The costs associated with making eye movements were minimal, given that the task was self-paced. Imposing a time limit or limiting the possible number of eye movements observers could make would be an interesting manipulation to see whether observers would make larger saccades and less fixations, which can be addressed by future research.

Our study shows evidence that peripheral vision is an important factor in mental maze-solving. The peripheral vision model not only predicts the difference in difficulty between wavy and straight mazes but also shows promise at predicting the distance along the maze path that an observer can perceive whether a point lies on or off the path. However, one cannot merely go directly from the peripheral vision model—nor from empirical measurements of the difficulty perceiving the path ahead using peripheral vision—to predicting the time and number of saccades required for an observer to solve a given maze. Rather, it is clear that one needs to understand decision processes and trade-offs involved in the visual system choosing the next fixation. Such trade-offs may, for instance, limit saccade length in maze-solving even when the observer can see farther along the path.

While the present study provides valuable insights into the role of peripheral vision in mental maze solving, it offers opportunities for future work. Notably, mazes used in these sets of experiments were simplified, featuring only a single enclosed path. Likewise, the manipulation of visual complexity was accomplished by merely altering the appearance of the maze walls in a simple and automatable way. This design choice was motivated by the intention to start with a simple approach before transitioning to more complex stimuli and tasks. Increasing cognitive, as well as visual, maze complexity by introducing additional branches or visual elements holds the potential to facilitate investigations into way-finding and backtracking behavior. Such extensions can enrich our comprehension of the involvement of peripheral vision in the context of maze-solving. It is our intent to address these facets through future research.

In summary, our findings make a substantive contribution to the existing body of literature pertaining to peripheral vision, visual crowding, and mental maze-solving. Many past studies of crowding have focused on particular peripheral tasks (e.g., peripheral object detection) or full-field tasks (e.g., scene perception). Much of the time, peripheral vision likely serves as one step in a multistep process, like using peripheral information to aid visual search ([Rosenholtz, Huang, & Ehinger, 2012](#); [Rosenholtz, Huang, Raj, et al.,](#)

2012) or find one's way through a maze. In previous investigations of mazes, we demonstrated that various perceptual attributes of maze design, including path length, thickness, and rendering of paths and walls, significantly impact maze-solving performance, thus implicating the role of crowding and visual complexity (Yu et al., 2019). Building upon this foundation, the present study combined eye tracking, modeling simulations, and targeted peripheral experiments to provide additional support to the notion that visual crowding significantly constrains the efficacy of mental maze-solving.

*Keywords:* peripheral vision, crowding, mental maze-solving

## Acknowledgments

Supported by grant NSF/NIH/BMBF IIS-1607486 to R. Rosenholtz, as part of the Collaborative Research in Computational Neuroscience Program.

Commercial relationships: none.

Corresponding author: Yelda Semizer.

Email: yelda.semizer@njit.edu.

Address: Department of Humanities and Social Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA.

## References

- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12), 13, doi:10.1167/9.12.13.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226, 177–178, doi:10.1038/226177a0.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Crowe, D. A., Averbeck, B. A., Chafee, M. V., Anderson, J. H., & Georgopoulos, A. P. (2000). Mental maze solving. *Journal of Cognitive Neuroscience*, 12(5), 813–827.
- Ehinger, K. A., & Rosenholtz, R. (2016). A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision*, 16(2), 13, doi:10.1167/16.2.13.
- Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6(9), 1, doi:10.1167/6.9.1.
- Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance while reading and while watching dynamic scenes* (pp. 269–293). Oxford, UK: Elsevier.
- Jolicoeur, P., & Ingleton, M. (1991). Size invariance in curve tracing. *Memory & Cognition*, 19, 21–36, doi:10.3758/BF03198493.
- Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, 14, 129–140, doi:10.3758/BF03198373.
- Jolicoeur, P., Ullman, S., & Mackay, M. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 997–1022.
- Keshvari, S., & Rosenholtz, R. (2016). Pooling of continuous features provides a unifying account of crowding. *Journal of Vision*, 16(3), 39, doi:10.1167/16.3.39.
- Korte, W. (1923). Über die gestaltauffassung im indirekten sehen [On the apprehension of Gestalt in indirect vision]. *Zeitschrift Fur Psychologie Mit Zeitschrift Fur Angewandte Psychologie*, 93, 17–82.
- Kowler, E., & Blaser, E. (1995). The accuracy and precision of saccades to small and large targets. *Vision Research*, 35(12), 1741–1754, doi:10.1016/0042-6989(94)00255-K.
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. chips: An ideal-observer model of reading. *Psychological Review*, 104, 524–553, doi:10.1037/0033-295X.104.3.524.
- Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: A mini-review. *Vision Research*, 48(5), 635–654, doi:10.1016/j.visres.2007.12.009.
- Najemnik, J., & Geisler, W. (2005). Optimal eye movement strategies in visual search. *Nature*, 434, 387–391, doi:10.1038/nature03390.
- Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research*, 49(10), 1286–1294, doi:10.1016/j.visres.2008.12.005.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46(28), 4646–4674, doi:10.1016/j.visres.2006.04.023.
- Pelli, D. G., & Tillman, K. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11, 1129–1135, doi:10.1038/nn.2187.
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7(2), 20, doi:10.1167/7.2.20.

- Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, 7(3), 6, doi:10.1167/7.3.6.
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuro Science*, 29, 203–227.
- Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science*, 2(1), 437–457, doi:10.1146/annurev-vision-082114-035733.
- Rosenholtz, R., Huang, J., & Ehinger, K. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, 3, 1–15, doi:10.3389/fpsyg.2012.00013.
- Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12(4), 14, doi:10.1167/12.4.14.
- Ullman, S. (1996). Visual cognition and visual routines. In *High-level vision: Object recognition and visual cognition* (pp. 263–1315). Cambridge, MA: MIT Press.
- Yu, D., Wan, Q., Balas, B., & Rosenholtz, R. (2019). Perceptual factors in mental maze solving. *Journal of Vision*, 19, 68, <https://doi.org/10.1167/19.10.68b>.
- Zhang, X., Huang, J., Yigit-Elliott, S., & Rosenholtz, R. (2015). Cube search, revisited. *Journal of Vision*, 15(3), 9, doi:10.1167/15.3.9.
- Zhao, M., Marquez, A. G., Hemmer, P., & Kowler, E. (2013). Inferring strategies of maze navigation from the movements of the eye and arm. *Journal of Vision*, 13, 124, doi:10.1167/13.9.124.