

Distributional and Temporal Properties of Eye Movement Trajectories in Scene Perception

Theo Rhodes (trhodes3@ucmerced.edu)
Christopher T. Kello (ckello@ucmerced.edu)
Bryan Kerster (bkerster@ucmerced.edu)

Cognitive and Information Sciences, University of California Merced, 5200 North Lake Rd., Merced, CA 95343

Abstract

Eye movements gather visual information from the environment for various purposes and goals. Spatial patterns of eye movements vary depending on the layout of visual information, and intentions of the observer. However, despite this variability, basic principles of visual information gathering may be reflected in lawful properties of eye movement trajectories that hold across various stimulus and intentional conditions. Two experiments are presented analyzing eye movement trajectories during scene perception across pictures with varying spatial frequency distributions (Expt 1), and across two different task conditions, "finding" versus "counting" tasks (Expt 2). Results show that, in all conditions, distributions of saccade amplitudes are heavy-tailed and nearly identical in shape, and fixation fluctuation series are long-range correlated with nearly identical spectral slopes. While a small effect of task intention was found, the broader conclusion is that eye movements during scene perception exhibit general statistical characteristics that models have yet to address.

Keywords: Eye movements, scene perception, lognormal distributions, Lévy flights, $1/f$ scaling, long-range correlation.

Introduction

Research on visual search and scene perception tends to focus on the effects of stimulus factors on eye movements. For instance, the debate over parallel versus serial search hinges on stimulus characteristics of targets, distractors, and the visual field (Triesman & Gelade, 1980). Models of scene perception relate the saliency of visual features and objects in scenes to probabilities of eye fixations (Itti, Koch & Niebur, 1998).

By contrast, the basic character of eye movements is mostly taken for granted in research on scene perception, i.e. there are saccades between fixations, and microsaccades and other more fine-grained movements within fixations (Liversedge & Findlay, 2000). These categories are coarse and describe little about the structure of eye movement trajectories, beyond the fact that trajectories will string together periods of small-scale movements (fixations) interspersed with periods of large-scale movements (saccades and pursuits).

One might assume that more quantitative statements about eye movements during scene perception will depend on particularities of scenes and intentions of observers. However, two bodies of research suggest otherwise. First, a large body of research on foraging behaviors has shown that search trajectories are nearly universally characterized by heavy-tailed distributions of path segment lengths (e.g.,

Sims et al., 2008). Various land and sea creatures have been tracked while foraging for food, and the lengths of paths from one locale to the next are measured. The probability of observing path length d often goes as $P(d) \sim d^{-\beta}$, with $\beta \sim 2$.

The precise formulation of the distribution is often a matter of dispute, but they are generally agreed to be heavy-tailed. The Lévy distribution is part of a broader class of heavy-tailed distributions that indicate multiplicative interactions in generating the observed data (i.e. path lengths in this case; Shlesinger, Zaslavsky & Klafter, 1993). These results are relevant to eye movements because animal foraging and scene perception are both search behaviors. Indeed, even memory search has been shown to result in heavy-tailed distributions of "path lengths", i.e. time intervals between recall events (Rhodes & Turvey, 2007).

These results suggest that eye movement trajectories may also exhibit heavy-tailed path length distributions by virtue of being a kind of search behavior. Consistent with this hypothesis, Stephen and Mirman (2010) found that distances between successive eye tracking samples were lognormally distributed in a "visual world paradigm" task (lognormal distributions are heavy-tailed and also associated with multiplicative interactions). This study alone, however, leaves it unclear whether the observed lognormal distributions were due to characteristics of the tasks or stimuli, such as their constrained, repetitive nature.

The second body of research to suggest general properties of eye movement trajectories concerns temporal correlations in neural and behavioral activity. It turns that many different measures of both kinds of activities have been found to contain long-range correlations in their intrinsic fluctuations (Kello et al., 2007). These correlations tend to follow a $1/f$ scaling relation, and $1/f$ scaling is also a kind of heavy-tailed distribution associated with multiplicative interactions (Van Orden, Holden & Turvey, 2003). Any time series can be expressed in the frequency domain as a set of sine waves of varying amplitudes (power) and frequencies (phase is discarded for this analysis). $1/f$ scaling describes a time series for which power is related to frequency as $P \sim 1/f^\alpha$, where ideally $\alpha \sim 1$.

Widespread findings of $1/f$ scaling, across modalities and levels of analysis, suggest that its origins are task-general and domain-general. $1/f$ scaling has also been found in eye movements, i.e. in fluctuations of repetitive target fixations (Shelhamer, 2005), and in variations within and across standard visual search tasks (Aks, Zelinsky & Sprott, 2002). However, as with heavy-tailed path lengths, these results on

1/f scaling in eye movements may also be particular to the repetitive, constrained nature of the tasks.

Scene Perception Experiments

The aim of our study was to investigate whether heavy-tailed distributions and long-range correlations are general facts about the statistical structure of eye movement trajectories. If so, then any theory or model of eye movements would need to explain how this structure is shaped to fit specific tasks and stimuli, while preserving its general character across conditions.

We chose to use scene perception as a general kind of visual information gathering that goes beyond previously used tasks. Most importantly, scene perception can unfold over relatively long periods of time without experimental intervention, and it embodies the kind of information gathering that visual systems are constantly engaged in.

We manipulated stimulus (Expt 1) and task (Expt 2) conditions in order to test whether findings are restricted to any particular conditions. Our basic dependent measure was Euclidean distances between adjacent pairs of eye tracking samples, which yields series of eye speed measurements. Speed series are appropriate for investigating general properties of eye movements because absolute position information is not preserved.

Experiment 1 Methods

Stimuli. Two hundred and fifty images were selected from a collection of National Geographic's Photo of the Day website. All images were originally 1600x1200 pixels in resolution, cropped to 1600x1100 to remove a watermark. Images were all natural scenes, color and spanned a wide range of content, including landscapes, action shots, close ups of animals and people.

To test whether the statistics of trajectories are affected by stimulus factors, we categorized them according to their spatial frequency distributions. Spectral analyses of spatial frequencies in natural images have been shown to exhibit 1/f scaling analogous to that in time series (Field, 1987). If heavy tails in eye movement trajectories come from visually searching over scenes with heavy-tailed spatial frequency distributions, then varying the latter should affect the former. Ten images were chosen within each of three categories: Steep ($\alpha < -3$), mid-range ($-3 < \alpha < -2.25$) and shallow ($\alpha > -2.25$) scaling relations.

Subjects, Apparatus, and Task. Eleven University of California Merced undergraduates participated in the experiment for course credit. Each participant was seated approximately 36" in front of a 30" flat panel LCD monitor. Participants viewed each of the thirty images in random order for 45 seconds per image. Monocular gaze position was recorded at 500 Hz using an Eye Link II head mounted eye tracker (Figure 1).

Subjects were instructed to view each image in the context of two tasks. One was a rating task whereby participants were asked after each image to characterize the complexity and memorability of the image on a scale of 1 to

10. The other task was a scene memory task whereby participants were given 60 seconds to verbally describe each of six images after each one was presented (because image order was randomized, the memory task appeared randomly to participants). Each subject viewed all 30 images.



Figure 1: Example visual search trajectory for an image of a spiral staircase inside ancient ruins.

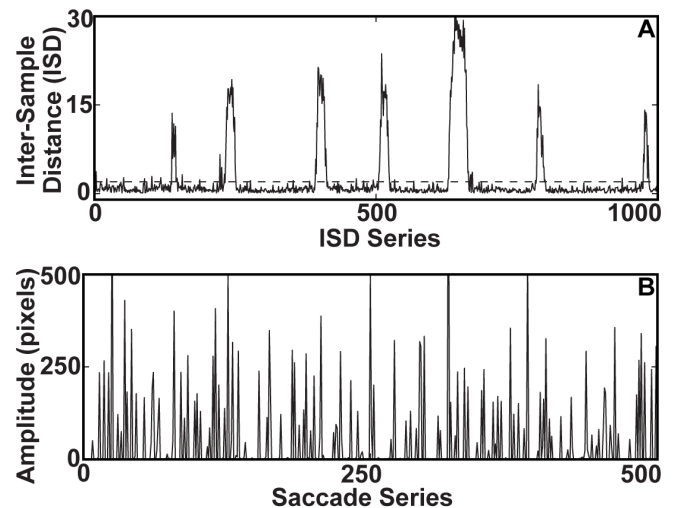


Figure 2: Example partial ISD time series (A) and saccade amplitude series (B). The dashed line in A indicates threshold, and B shows the lengths above threshold.

Experiment 1 Results

From the raw gaze position series (Figure 1), inter-sample distances (ISDs) were computed (Figure 2) and blinks and measurement malfunctions were removed. As one would expect given the well known fixation and saccade structure of human gaze trajectories (Henderson, 2003), ISD series were characterized by clusters of low values (fixations) interspersed with “bursts” of high values (saccades). Based on visual inspection, a threshold of 1.5 pixels was used to separate saccade speed “bursts” from fixation fluctuations (increasing this threshold by a few pixels had no qualitative effect on results). Each saccade length (i.e. amplitude) was

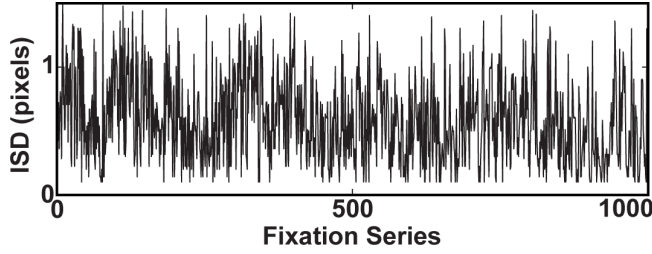


Figure 3: Example below-threshold fixation time series.

measured as the distance from the first point over threshold to the next point below threshold (Figure 2), and ISDs below threshold were concatenated to yield fixation fluctuations.

Saccade amplitudes correspond to the path lengths of a Lévy flight, so their distributions were examined. Fixation ISDs corresponded to intrinsic fluctuations in speed (with saccades removed), in that no experimental manipulations were applied within trial series, i.e. for each given picture (note that speed may be affected by the location of gaze on a given scene, but the experimenter does not directly influence where to gaze). Thus their temporal correlations were examined.

Figure 4 shows saccade amplitude distributions in log-log coordinates for steep, mid-range, and shallow stimulus conditions, aggregated over all trials and subjects. Multi-model inference (Burnham & Anderson, 2002; Wagenmakers & Farrell, 2004) was used to determine the statistical model most likely to generate these distributions. Four candidate statistical models were tested: exponential, lognormal, gamma and Pareto distributions. Lognormal, gamma and Pareto distributions are heavy-tailed, suggestive of a complex, multiplicative process. The lognormal distribution is one of the simplest multiplicative distributions, the Pareto distribution represents an idealized power-law function, and the gamma distribution is a hybrid of power-law and exponential distributions, indicating a truncated power-law. The lognormal distribution has all power moments defined while the Pareto and gamma distributions do not. For each candidate, the negative log-likelihood (the function maximized in maximum likelihood estimation) was calculated. Log-likelihoods were then used to calculate Akaike's information criterion, which is a measure of the information-theoretic distance between candidate distributions and the distribution of the data. The minimum AIC value indicates the most likely candidate distribution given the data.

AIC strongly supported the lognormal distribution for all three condition aggregates. Aggregates were also created for each image and each subject, and lognormal was supported for 100% of the former, and 100% of the latter. Using maximum likelihood estimation, we calculated the parameters corresponding to the best fit lognormal distribution. There were no significant effects of condition, image or their interaction (mean $\sigma = 1.25$). AIC analysis of individual trials predominantly supported lognormal (60%) with some support for gamma (38%) and pareto (2%)

distributions. Thus there was absolutely no support for the exponential distribution without a heavy tail.

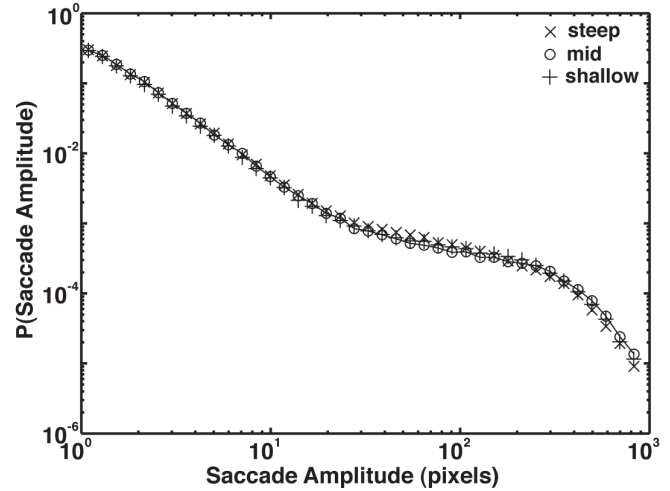


Figure 4: Aggregate saccade amplitude distributions from Experiment 1.

Spectral analyses of fixation amplitude series resulted in evidence for long-range correlations. Averaged spectra for each condition are shown in Figure 5 in log-log coordinates. $1/f$ scaling appears as a negative linear relation in log-log spectra, with a slope significantly less than zero. The figure shows a $1/f$ scaling relation in the lower frequencies (where most variation resides, despite appearances due to the logarithmic scales), but it is difficult to distinguish long-range from short-range correlation by visual inspection (Wagenmakers, Farrell & Ratcliff, 2004).

To distinguish between short-range and long-range correlation, we applied a maximum likelihood method developed by Thornton and Gilden (2005; Torre & Wagenmakers, 2009).

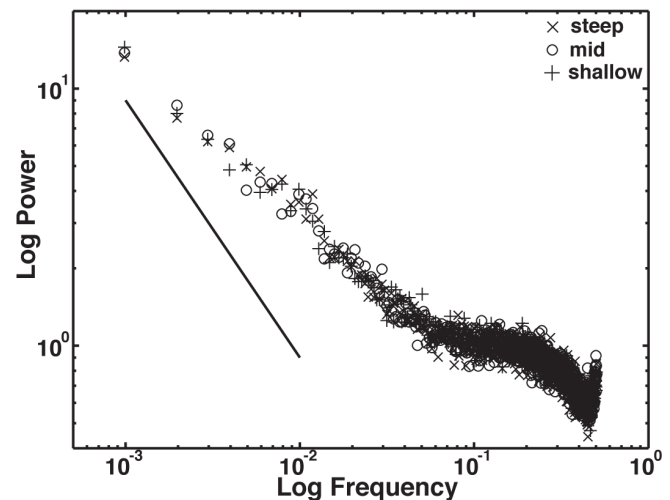


Figure 5: Averaged spectra for fixation fluctuation series in Experiment 1 (dark line is $\alpha = 1$ reference point).

The method differentiates between two qualitatively different models of time series data. One is a $1/f$ scaling model, which they expressed in terms of fractional Brownian motion plus white noise (fBmW). The other model is short-range correlation model, which they express as an autoregressive moving-average (ARMA), also with a white noise term. Both models are thus specified by two parameters, removing potential bias towards a model with more free parameters. Given a predefined parameter space, the maximum likelihood method chooses the model and pair of parameter values most likely to generate fluctuations with the spectral characteristics of an observed fluctuation series.

A spectrum was computed for each of the individual peak angle series, and averaged to form equivalent aggregates for each condition, each image, and each subject. In all cases, fBmW was supported over the ARMA model. The α parameter estimates were also compared across conditions, images, and subjects. There was no significant effect of condition (mean $\alpha = 0.60$, $F(2,327) = 0.48$).

Experiment 1 Discussion

These results provide broader evidence that heavy-tailed distributions and long-range correlations are general properties of eye movement trajectories. Previous studies found heavy-tailed distributions for restricted sets of artificial stimuli, under highly constrained task conditions, and also long-range correlations in times to repeatedly saccade between pairs of targets (Shelhamer, 2005). Experiment 1 expanded upon these results using a wider range of natural stimuli, a more natural scene perception task, and a novel but simple threshold method for dividing series of eye movement speeds into saccade amplitudes and fixation fluctuations (above and below threshold).

The AIC analysis of individual trials shows majority support either for lognormal or gamma distributions. The gamma distribution is sometimes characterized as a truncated power law, as it is essentially a power law distribution that tapers into an exponential distribution. Lognormal distributions are also common in situations where heavy-tailed behavior is bounded, either in time or magnitude. Given that eye-tracking is sharply bounded by the dimensions of the screen, support for lognormal and gamma may be convergent support for a truncated power law (a task for future research).

While heavy-tailed distributions and $1/f$ scaling held across images and participants, the task in Experiment 1 was homogeneous across trials. The effects of intention on scene perception are well-known and were first documented by Yarbus (1967), who showed that trajectories over the same image were noticeably different depending on viewing instructions. Yarbus' work was limited in scope, but more recent replications have lent some quantitative support to his results (DeAngelus & Pelz, 2009). Thus different intentions may result in different distributions and temporal properties.

Experiment 2 was designed to test this possibility. We created two scene perception search tasks with the goal of generating highly distinct eye movement trajectories. In a

“find” task, participants searched for a small star embedded in the image. In a “count” task, participants searched for all objects of a given kind in order to count them.



Figure 6: Example visual search trajectory for the find task (top) and counting items task (e.g. sheep; bottom).

Experiment 2 Methods

Stimuli. Images were selected from the same collection as Experiment 1 with the additional criteria that they contain a set of 40-100 enumerable objects. Example sets include sheep, spots on a giraffe, and leaves floating in a lake. A total of 30 such images were selected. A second version of each image was created by embedding a small, transparent and textured star at a difficult-to-find location in the image. A group of volunteers provided feedback to calibrate the difficulty of locating stars, and star locations and transparencies were adjusted such that stars could be found with 60 seconds approximately 50% of the time.

Subjects, Apparatus, and Task. Sixteen University of California Merced undergraduates participated in the experiment for course credit. Participants were carefully screened for potential issues that might affect sampling rate, such as eyeglasses or cosmetics. Each participant was seated approximately 36" in front of a 30" flat panel LCD monitor. Participants viewed one set of 15 images in random order,

and were asked to either find the star, or count the instances of a given object. They then performed the other task on the other 15 images, also presented in random order. Task and set order were counterbalanced, with four participants in each of the four possible combinations. Participants were given 45 seconds for each image, and in the case of the find task, were instructed to fixate and press a key should they locate the star before time was up. Monocular gaze position was recorded at 500 Hz using an Eye Link II head mounted eye tracker.

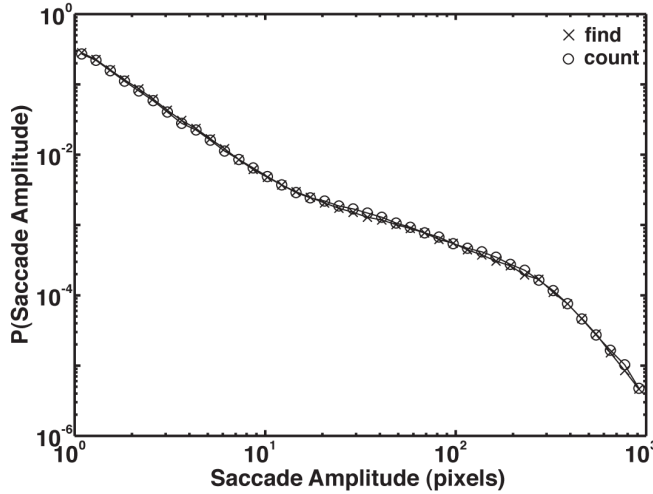


Figure 7: Aggregate saccade amplitude distributions from Experiment 2.

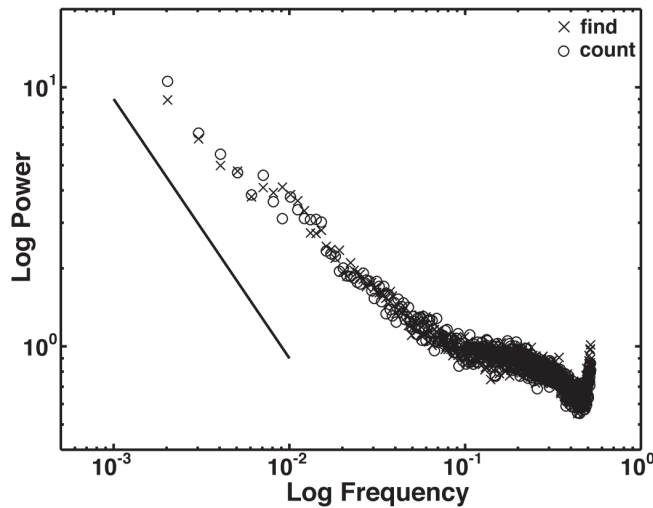


Figure 8: Averaged spectra for fixation fluctuation series in Experiment 2 (dark line is $\alpha = 1$ reference point).

Experiment 2 Results

The different task conditions had the desired effect of evoking clearly distinct spatial patterns of eye fixations. Figure 6 shows example trajectories for the find task (upper) versus the count task (lower). Based on visual inspection, the large majority of find trials were similarly distinct from the large majority of count trials. This means that positional

information of eye movement trajectories is distinct (a topic for future research), but eye movement speeds do not carry positional information and may thus exhibit the same general properties found in Experiment 1.

Saccade amplitudes and fixation fluctuations were constructed as in Experiment 1, and both series exhibited the same properties as in Experiment 1. Support for lognormally distributed saccade amplitudes was strong at both the aggregate and individual trial levels (Figure 7). In particular, lognormal was supported for 86% of the find trials, 94% of the count trials, 100% of the participant aggregates, and 100% of the image aggregates, with the remaining trials in both cases showing support for a gamma distribution. Estimates of the lognormal shape parameter were slightly but reliably lower for the find task trials (mean $\sigma = 1.25$) compared with the count task trials (mean $\sigma = 1.26$, $p < 0.001$).

Also as in Experiment 1, spectral analysis was applied to each series of fixation ISDs from each trial. Spectra were averaged for find trials versus count trials (Figure 8), and the Thornton and Gilden (2005) method was used to test for long-range correlations (versus short-range or no correlations). We found strong evidence for the fBmW model in aggregate spectra for each condition, image, and subject. In all cases, fBmW was supported over the ARMA model. There were no significant differences in the long-range correlation parameter estimates ($\alpha = 0.80$) between task conditions.

Experiment 2 Discussion

The results of Experiment 2 were as unequivocal as Experiment 1: the properties of heavy-tailed distributions and long-range correlations remained, despite visible differences in the spatial layouts of eye movement trajectories. Task condition had a small but reliable effect on lognormal parameter estimates for saccade amplitude distributions, which indicates that task intentions can alter series of eye movement speeds, at least by slight quantities. There were also differences in lognormal and 1/f parameter estimates across experiments that appear to be at least partly due to methodological differences. The fact that results were otherwise so uniform across experiments provides further evidence for their generality.

Conclusions

The finding that eye movement speeds during scene perception have a common statistical structure may not seem very interesting at first, at least not to a cognitive scientist. All humans have oculomotor apparatus and control systems, so one might expect these similarities to result in common statistics of measures like speed that do not carry positional information. The present results, however, go well beyond typical measures of central tendency and variance. Series of eye movement speeds were found to contain two different general properties in their saccade amplitudes versus fixation fluctuations. Both of these properties are power laws, at least within a given range of

scales. And both have been found in other studies of eye movements, as well as other studies of neural and behavioral activities of many different kinds. Thus heavy-tailed distributions and long-range correlations appear to be common to eye movement trajectories.

The implications of these results are yet to be explored. One might be able to formulate theories of search behaviors that provide a common basis for understanding searches through habitats, visual fields, information networks (e.g. world-wide web), and long-term memory (Hills, 2006). For instance, current directions in foraging research include complex diffusion and state-space models based on the boundary conditions and constraints specific to search environments (Patterson et al., 2008). Similar models may prove fruitful in understanding visual search, especially based on models of the environment such as saliency maps.

It is important to note that the above mentioned search models are typically aimed at explaining path lengths (saccade amplitudes) as opposed to fixation fluctuations. With regard to the latter, long-range correlations spanned the many dozens of fixations (interspersed by saccades) that occur in 45 seconds of scene viewing. This finding is in the purview of scene perception models because its time scale goes well beyond “low-level” mechanisms like image stabilization, at least as they are currently formulated. Thus it will be a challenge to formalize models that guide eye movements over varied scenes for varied purposes of information foraging, while also generating long-range correlations in the fluctuations of eye movement speeds.

With regard to empirical directions, further statistical commonalities may be found in the spatial distributions of eye fixations that are more typically the focus of scene perception models (Henderson, 2003), and appear to be more greatly affected by stimulus and task factors. Whatever the case, it will be informative to investigate how models and theories of scene perception might address the present results.

Acknowledgments

The authors would like to thank Michael Spivey and Teenie Matlock for their helpful contributions. The work was funded by NSF grants 08-42784 and 10-31903.

References

- Aks, D. J., Zelinsky, G., & Sprott, J. C. (2002). Memory across eye- movements: 1/f dynamic in visual search. *Journal of Nonlinear Dynamics & the Life Sciences*, 6, 1–25.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection, and multimodel inference: A practical information-theoretic approach*. New York: Springer.
- Castelhano, M. S., Mack, M. L. & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9, 1-15.
- DeAngelus, M. & Pelz, J. B. (2009). Top-down control of eye movements: Yabus revisited. *Visual Cognition*, 17, 790-811.
- Field, D. J. (1987). Relations between the statistics of natural images and the response profiles of cortical cells. *Journal of the Optical Society A*, 4, 2379–2394.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7, 498-504.
- Hills, T. T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive Science*, 30, 3-41.
- Itti, L., Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254-1259.
- Kello, C. T., Beltz, B., Holden, J. G. & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General*, 136, 551-568.
- Liversedge, S. P. & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4, 6-14.
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology and Evolution*, 23, 87-94.
- Rhodes, T. & Turvey, M. T. T. (2007). Human memory retrieval as Lévy foraging. *Physica A*, 385, 255-260.
- Shelhamer, M. (2005). Sequences of predictive saccades are correlated over a span of ~2s and produce a fractal time series. *Journal of Neurophysiology*, 93, 2002-2011.
- Shlesinger, M. F., Zaslavsky, G. M. & Klafter, J. (1993). Strange kinetics. *Nature*, 363, 31-37.
- Sims, D. W., Southall, E. J., Humphries, N. E., Hays, G. C., Bradshaw, C. J. A., Pitchford, J. W., et al. (2008). Scaling laws of marine predator search behavior. *Nature*, 451, 1098–1102.
- Stephen, D. & Mirman, D. (2010). Interactions dominate the dynamics of visual cognition. *Cognition*, 115, 154-165.
- Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin & Review*, 12, 409-441.
- Torre, K. & Wagenmakers, E.-J. (2009). Theories and models of 1/f^β noise in human movement science. *Human Movement Science*, 28, 297-318.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132, 331–350.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11, 192-196.
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/f noise in human cognition. *Psychonomic Bulletin & Review*, 11, 579–615.
- Yarbus, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York: Plenum Press.