Multi-Scale Entropy (MSE) of Eye Movements

by

Vinoja Charathchandra Sebanayagam

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Institute of Medical Science
University of Toronto

© Copyright by Vinoja Charathchandra Sebanayagam 2018
Multi-Scale Entropy (MSE) of Eye Movements

Vinoja Charathchandra Sebanayagam

Master of Science

Institute of Medical Science
University of Toronto

2018

Abstract

Conventional eye movement (EM) metrics are static in that they quantify only one particular aspect of vision (e.g. number of gaze fixations made in a trial etc.). Therefore, in the current study we use a metric called Multi-Scale entropy (MSE) to quantify the spatial and temporal changes in EMs as viewing unfolds over time. In particular, we aimed to capture the inherent complexity and information contained in gaze signals produced by EMs when viewing different types of scenes. Our results showed that MSE can be applied on gaze signals and that it does quantify the meaningful information contained therein. It also indexed differences in the information encoded from different scenes beyond what can be deduced from conventional metrics like the number of fixations and saccades. However, we also found that this metric is related to and is comparable with saccade amplitude with respect to indexing scene differences.
Acknowledgments

Graduate school has been a unique experience that helped me grow both personally and professionally. It helped me improve many skills including leadership, the ability to collaborate with others, communicate ideas and thoughts effectively, and the ability to ask questions and find answers. It also reinforced my desire to be a lifelong learner.

This particular thesis has been in the making for a long time and I have crossed a few hurdles to get to this point. I would first like to thank Dr. Jennifer Ryan, who has been my primary supervisor and research mentor since my 2nd undergraduate year. Dr. Ryan not only introduced the world of research to me and constantly supported me in my graduate school journey, but also helped me pursue my lifelong ambition of becoming a medical doctor. Similarly, I would like to thank Dr. Randy McIntosh for also being extremely supportive of my professional ambitions and my graduate work. I would also like to thank Dr. Cheryl Grady and Dr. Morgan Barense for their guidance and support since I started my project in 2015. I would especially like to thank Dr. Kelly Shen for providing me a solid factual and computational understanding of this project. I would not have the programming and theoretical knowledge that I have now, if not for her constant support and guidance throughout this project.

Moreover, I would like to thank my parents for being exceptional role models to me, both personally and professionally. I would not be able to pursue my dreams today if not for their hard work and sacrifices. I am also extremely thankful to all my lab mates for supporting and encouraging me during graduate school. Finally, I thank all my friends for providing moral support throughout all these years and for always being there for me when I needed them.
Contributions

1. Dr. Kelly Shen provided suggestions, instructions and feedback with the computational analyses involved in devising the metric described in this thesis. She also gave me feedback on initial drafts of this thesis.

2. Arber Kacollja contributed to subject recruitment, eye movement data collection and preparation.

3. The scene stimuli used in this study were taken from a public scene dataset cited in this thesis.

The citation, as suggested by authors of this site: Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. MIT Saliency Benchmark. URL: http://saliency.mit.edu/
# Table of Contents

Acknowledgments.................................................................................................................. iii

Contributions........................................................................................................................ iv

Table of Contents.................................................................................................................... v

List of Tables ........................................................................................................................... viii

List of Figures ......................................................................................................................... ix

List of Equations ..................................................................................................................... xi

Chapter 1 .................................................................................................................................... 1

1 Literature Review ..................................................................................................................... 1

1.1 Introduction ......................................................................................................................... 1

1.1.1 Bottom-up control of eye movements ........................................................................... 2

1.1.2 Top-down control of eye movements ......................................................................... 3

1.2 Relationship between eye movements, perception and cognition ....................................... 5

1.2.1 Role played by eye movements in the cognitive processing of information (behavioral evidence) .......................................................... 6

1.2.2 Structural and functional evidence for the relationship between eye movements and cognitive processes (focus: memory) ....................... 7

1.2.3 Eye movement monitoring has revealed individual differences in the association between eye movements and perceptual/cognitive processes .... 8

1.3 Static measures of eye movements ....................................................................................... 9

1.4 Non-linear Dynamic System (NDS) .................................................................................... 11

1.5 Characterizing the dynamic nature of perception and cognition using a NDS approach .. 13

1.6 Characterizing the dynamic nature of eye movements using a NDS approach ................. 15

1.6.1 Overview of the nature and dynamics of eye movements ............................................ 16

1.7 Quantifying the temporal and spatial dynamics of eye movements .................................. 17

1.7.1 Effective dimensions of power laws that characterize the scaling behavior of eye movements ................................................................. 18
1.7.2 Temporal dynamics of eye movements ............................................................ 19
1.7.3 Spatial dynamics of eye movements ............................................................. 21
1.8 Importance of quantifying spatial and temporal dynamic properties of eye movements .. 23
1.9 Relationship between signal regularity, entropy and information .......................... 25
1.10 Usage of Shannon’s entropy and its derivatives .................................................. 26
   1.10.1 Topological entropy (Kolmogorov-Sinai entropy) ........................................ 26
   1.10.2 Approximate entropy measures and sample entropy ................................... 27
   1.10.3 Zhang’s entropy ......................................................................................... 28
1.11 Multi-Scale Entropy (MSE) ................................................................................. 29
1.12 Previous applications of MSE ............................................................................ 31
   1.12.1 Application on cardiac inter-beat interval timeseries .................................... 32
   1.12.2 Application on coding and non-coding sequences of DNA ......................... 32
   1.12.3 Application on neural signals ........................................................................ 33
1.13 Current attempt: application of MSE as an eye movement complexity measure ......... 35
1.14 Overview of thesis ............................................................................................ 36
Chapter 2 ................................................................................................................. 39
2 Research aims and hypotheses ............................................................................... 39
Chapter 3 ................................................................................................................. 45
3 Experimental study .............................................................................................. 45
   3.1 Introduction ...................................................................................................... 45
   3.2 General materials and methods ....................................................................... 52
      3.2.1 Participants .................................................................................................. 52
      3.2.2 Raw eye movement recordings ................................................................... 52
      3.2.3 Analysis of Visual Sampling Behavior ...................................................... 56
      3.2.4 Eye movement data reduction (calculation of angular displacement) .......... 57
      3.2.5 Fast Fourier Transform analysis ................................................................. 57
List of Tables

Table 1 – The table depicts visual sampling behavior for each scene category..................65
List of Figures

Figure 1. Sample images taken from the 12 scene categories selected for the current experiment are depicted in this figure………………………………………………………………………………..55

Figure 2 - The top panel illustrates the nature of gaze angular displacement as a function of time, calculated for a sample trial (an art scene viewed by a participant) and captured at finer (timescale = 1, 30) and coarser (timescale = 60) timescales. The bottom panel depicts sample entropy calculated from angular displacement signal patterns that correspond to timescales 1, 30 and 60…………………………………………………………………………………………….61

Figure 3 – (a) The figure depicts a power spectrum of signal power versus frequency, obtained from a discrete Fourier Transform (DFT) applied on gaze displacement signals derived from a subject and trial (art scene). (b) The plot depicts a power spectrum of log transformation of the power spectrum (log-log plot of power versus frequency). (c) The plot depicts the linear high frequency signal region (log (frequency) > 1) of the log-log plot of the power spectrum………67

Figure 4 – The figure depicts the average MSE (+/- SE) profiles for each scene category: (a) Art, (b) Fractal, (c) Inverted, (d) Jumbled, (e) Pattern, (f) Random, (g) Action, (h) Cartoon, (i) Indoor, (j) Outdoor Man-made, (k) Outdoor Natural and (l) Social……………………………..68

Figure 5 – Average MSE profiles of all 12 scene categories are presented in the same plot in this figure……………………………………………………………………………………………………69

Figure 6 – (a) The plot depicts the data-driven contrast between scene category specific MSE values that was significantly expressed across the timescales, as determined by permutation tests. (b) The plot depicts the bootstrap distribution, which highlights the temporal (timescale-based) reliability of the contrast depicted in figure 6(a), across all subjects……………………………..70
Figure 7 – (a) The plot depicts the data-driven contrast in residuals between scene categories that was significantly expressed across the timescales, as determined by permutation tests. The number of saccades was regressed from MSE to produce the residuals. (b) The plot depicts the bootstrap distribution, which highlights the temporal (timescale-based) reliability of the contrast depicted in figure 7(a), across all subjects………………………………………………………………………72

Figure 8 – (a) The plot depicts the data-driven contrast in residuals between scene categories that was significantly expressed across the timescales, as determined by permutation tests. The average saccade amplitude was regressed from MSE to produce these residuals. (b) The plot depicts the bootstrap distribution, which highlights the temporal (timescale-based) reliability of the contrast depicted in figure 8(a), across all subjects……………………………………………………......73

Figure 9 – The plot depicts a significant positive correlation between the peak MSE, and saccade amplitude, both represented by trial, across the 12 scene categories………………………………………………..74

Figure 10 – The plot shows a significant negative correlation between peak MSE and average NSS score across the 12 scene categories…………………………………………………………………..75

Figure 11 – The plot depicts a significant negative correlation between the average saccade amplitude measure and average NSS score, both represented by scene category………………………………………76
List of Equations

Equation 1 – Calculation of each element of a given coarse-grained timeseries.………………59

Equation 2 – Calculation of sample entropy at each timescale.……………………………………60
Chapter 1

1 Literature Review

1.1 Introduction

The current thesis is a novel attempt to apply a metric called Multi-scale entropy (MSE) on gaze signals derived from eye movements, in order to quantify the complexity of these signals, and as an extension, the dynamics of eye movements. In particular, we have quantified the complexity of gaze signals obtained from viewing different types of scenes and compared MSE to basic eye movement metrics that are currently used in the field of vision research.

During viewing, ballistic eye movements called saccades enable the eye to foveate (make gaze fixations) on different regions of a stimulus to obtain information from the stimulus. In this manner, saccades contribute to the accumulation of visual information from the external viewing space (Irwin, 1991; Pertzov, Avidan, & Zohary, 2009; Rayner & Pollatsek, 1992). The information accumulated by saccades is then used to build a stable and continuous perception of the world (Melcher & Morrone, 2007). We selectively attend to specific stimuli from the large amount of visual input that reaches our retinas, which aids in subsequent perceptual processing of those stimuli, and facilitates cognitive and/or motor functions (Borji & Itti, 2015; Borji, Sihite, & Itti, 2013). Evidence for such selective information gathering behavior has been put forward by several researchers in the past. For instance, Buswell (1935) showed that during scene viewing, individuals tend to allocate a large proportion of gaze fixations to a small portion of the scene, as opposed to randomly making fixations over the entire picture, suggesting that gaze is
targeted on informative areas of a given picture (Antes, 1974; Henderson, Weeks, Phillip, & Hollingworth, 1999; Loftus & Mackworth, 1978).

Eye movements that aid in selective visual exploration are guided by factors that can be broadly categorized as bottom-up vs top-down in nature (Antes, 1974; Henderson, 2003; Henderson et al., 1999; Loftus & Mackworth, 1978; Parker, 1978). Research done by Henderson and colleagues (1999) suggests that eye movements are influenced by bottom-up factors before they are influenced by top-down control mechanisms during the time course of viewing. These factors contribute to two types of information contained in scenes: bottom-up information, which is largely based on the physical features of stimuli, and top-down information, which relates to stimuli or parts of stimuli that are cognitively relevant to observers (Loftus & Mackworth, 1978).

1.1.1 Bottom-up control of eye movements

Bottom-up stimulus features guide eye movements in specific ways that enable the gathering of information from these features. Informal evidence suggests that individuals tend to make more fixations on regions of a scene that show physical discontinuity such as unpredictable contours (Mackworth & Morandi, 1967). Similarly, individuals allocate the majority of fixations to the less redundant (e.g. involving less repetition) of two simple visual stimuli when redundancy is defined in physical terms, such as a regular vs irregular checkerboard (Berlyne, 1958). These studies suggest that bottom-up information obtained via viewing can be attributed to eye movements (e.g. specific pattern of fixations) that are guided by intra-stimulus redundancy based on physical features of stimuli (Loftus & Mackworth, 1978). Other image properties that increase the number of fixations made on certain regions of a scene include high spatial frequency content, edge density (Mannan, Wooding, & Ruddock, 1996; Mannan, Ruddock, & Wooding, 1997), and local contrast, as well as low two-point correlation (i.e. intensity of the
fixated point and near-by points; Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Reinagel & Zador, 1999; Henderson, 2003). However, it has been observed that beyond the initial saccades that might be dictated by bottom-up factors, eye movements are largely controlled by top-down influences (Henderson et al., 1999) such as cognitively/semantically relevant parts of stimuli (Loftus & Mackworth, 1978).

1.1.2 Top-down control of eye movements

Top-down control of eye movements relates to the semantic/cognitive information contained in visual stimuli or their cognitive relevance to the observer that can guide eye movements. The nature and extent of cognitive information contained in visual stimuli (e.g. complex scenes) have been characterized in previous studies through different measures. For instance, in the research done by Loftus & Mackworth (1978), an object in a picture was considered to be informative, from a cognitive perspective, if it had a low a priori probability of being present in the picture given the context of picture and the observer’s past history. Cognitively informative parts of stimuli guide eye movements in specific ways as shown by observers making significantly more fixations on objects whose presence was somewhat surprising based on the semantic context of the picture (e.g., viewing an octopus in a barnyard scene; Loftus & Mackworth, 1978; Yarbus, 1967). Moreover, individuals also tend to make fixations earlier and over longer durations when viewing objects that have been moved from a previously viewed location, or are semantically anomalous, or positioned within a semantically anomalous location, in a given scene (Henderson et al., 1999; Loftus & Mackworth, 1978).

While the above studies have used how predictable an object in a scene is to gauge the cognitive information content of scenes, other studies have gauged this using the future ease of recognition of certain sections of complex scenes (Antes, 1974; Mackworth & Morandi, 1967). These studies
have shown that individuals concentrate their eye fixations on regions of pictures that were subsequently rated as highly recognizable (hence, informative; Mackworth & Morandi, 1967). However, the time course of eye movements differed based on whether the cognitive information ratings were obtained from independent subjects (from whom eye movement data was not obtained, i.e. in this case, eye movement data was obtained from a different group of subjects) compared to subjects from whom eye movement data was also later obtained while they viewed the pictures. Even among studies that recorded eye movements of the same subjects (while they viewed the pictures) who rated certain regions of the picture as highly informative prior to viewing, there seems to be a discrepancy in the time course of viewing. For example, Mackworth & Morandi (1967) showed that individuals fixated on informative regions of the scene immediately and continued to do so over the entire viewing period. However, Buswell (1935) and Karpov, Luria, & Yarbus (1968), who used the same method as Mackworth & Morandi (1967; eye movements were recorded from the same subjects who rated certain sections of pictures as informative prior to the eye movement monitoring session), suggested that there might be an initial orientation period during which the entire picture is surveyed, followed by a focus on highly informative regions of the picture. On the other hand, when the cognitive information content of pictures was rated by independent subjects (different from those whose eye movements were recorded while viewing the pictures) in a study done by Antes (1974), researchers observed that the information content of locations fixated by individuals decreased over time and reached a plateau after 10 sec of viewing time. This study indicated that as a response to the cognitive information content of pictures, people made many long saccades initially to fixate on informative elements of the picture for short (fixation) durations. This behavior gradually evolved into fixating on informative elements less frequently and examining less informative elements for longer periods over the course of viewing. However, regardless of
the manner in which cognitive information content of a picture is scaled, the above studies collectively suggest that semantic/cognitive information contained in pictures imposes top-down control on the nature and progression of eye movements, which consequently predicts the information acquired through viewing.

1.2 Relationship between eye movements, perception and cognition

During stimulus (e.g. scene) perception, high quality information is obtained from the region that immediately surrounds the center of the gaze (foveated by the eye; Henderson, 2003). Through saccadic eye movements, we reorient the fovea over a given stimulus to gather information (Henderson, 2003). However, the information required to perceive stimuli is only acquired during fixations where the gaze remains stable due to saccadic suppression (Matin, 1974; Rayner, 1998; Thiele, Henning, Kubischik, & Hoffmann, 2002). This process of gaze control helps direct fixations to informative areas of stimuli in real time, to serve ongoing perceptual processing of information (Henderson, 2003).

Eye movements have been used to indirectly index individual differences in perceptual and cognitive processes in various fields, including cognitive and behavioral neuroscience, psychology and psychiatry. With respect to perceptual processing, eye movements are useful in acquiring selected information from the large amount of visual stimulation that constantly hits our retinas, to subsequently support the perception of stimuli (Borji & Itti, 2013, 2015; Henderson, 2003; Itti, Koch, & Niebur, 1998). Eye movements are also known to reveal the contents of cognitive processes, including memory, that cannot always be accurately quantified by explicit behavioral tests (Althoff & Cohen, 1999; Ryan, Althoff, Whitlow, & Cohen, 2000).
1.2.1 Role played by eye movements in the cognitive processing of information (behavioral evidence)

Previous research has defined specific eye movement-based phenomena that reflect cognitive processes, particularly memory. These phenomena include the eye movement repetition effect (Althoff & Cohen, 1999; Heisz & Ryan, 2011; Olsen et al., 2016; Smith & Squire, 2017) and relational manipulation effect (Ryan et al., 2000). The eye movement repetition effect involves a reduction in sampling behavior of previously viewed stimuli compared to novel stimuli. This reduction is quantified by a decrease in the number of eye fixations made or regions sampled, indicating memory for those stimuli (Althoff & Cohen, 1999; Heisz & Ryan, 2011; Olsen et al., 2016; Ryan et al., 2000). Moreover, viewing is thought to become somewhat deterministic with repeated stimulus exposure in that it involves less random transitions (saccades) between successive fixations during repeated versus initial viewing (Yarbus, 1971; Ryan et al., 2000).

Similarly, the relational manipulation effect involves an increase in viewing of regions of a scene that have undergone some form of manipulation (e.g. removing an element from the scene) after initial viewing (Ryan et al., 2000). This effect has been observed in individuals who were not aware of a scene manipulation, suggesting that eye movements can reveal information about spatial relations among scene elements contained in memory representations of scenes, even when they are unavailable for conscious (verbal) report (Ryan et al., 2000).

While the two effects mentioned above relate to how eye movements reflect memory for previously viewed items or relations among stimulus (e.g. scene) elements, other studies have shown the influence of eye movements on memory itself. These studies show that eye movements enacted during stimulus viewing can actively influence/inform memory processes, including memory encoding and retrieval. For instance, eye fixations made during stimulus
encoding/learning have been linked to the acquisition of subsequent memory (Chan, Kamino, Binns, & Ryan, 2011; Henderson, Williams, & Falk, 2005; Olsen et al., 2016). Similarly, research suggests that eye movements occurring during memory retrieval might serve as facilitatory cues for retrieval, indicating the active role played by eye movements in this memory process (Norton & Starks, 1971; Johansson, Holsanova, Johansson, Dewhurst, & Holmqvist, 2012; Johansson & Johansson, 2014; Wynn et al., 2016). Meanwhile, other studies have suggested that eye movements might be indicators or markers of retrieval in that they merely express (reflect) the memory process (Hannula & Ranganath, 2009; Ryals, Wang, Polnaszek, & Voss, 2015).

1.2.2 Structural and functional evidence for the relationship between eye movements and cognitive processes (focus: memory)

Recently, the behavioral link that has been established between eye movements and memory from past research gained further support from anatomical (structural) and neuroimaging (functional) research. By exploring the directed anatomical connectivity in the macaque brain using a data driven network-based approach, Shen et al., 2016 revealed an extensive set of polysynaptic pathways across the extra-striate, posterior parietal and prefrontal cortices that have the potential to mediate information exchange between the oculomotor (supporting vision) and memory systems. It is also now known that regions that are involved in the cognitive control of saccades, such as the dorsolateral prefrontal cortex and the frontal eye field (part of the oculomotor control system) are topologically well positioned in a manner that supports directed information flow from the hippocampus (supports memory) to the oculomotor system (Shen, Bezgin, Selvam, McIntosh, & Ryan, 2016). This suggests that the structural link between the two systems might have functional relevance.
Subsequently, Liu et al. (2017) explored the functional connection between the visual and memory systems by directly analyzing the relationship between gaze fixations and activation in the hippocampus. They found that the gaze fixations made by subjects during face viewing was significantly correlated with hippocampal activity. In particular, this association was seen when subjects viewed novel faces relative to repeated faces, indicating that this association might be a reflection of the ongoing development of memory representations (Liu, Shen, Olsen, & Ryan, 2017). This study suggested that visual exploration might be functionally linked to hippocampal binding processes that support memory formation (Liu et al., 2017). Hence, converging evidence suggests that eye movements could serve to both facilitate (e.g. memory encoding and retrieval) and express (e.g. repetition effect, relational manipulation effect and memory retrieval) cognitive processes.

1.2.3 Eye movement monitoring has revealed individual differences in the association between eye movements and perceptual/cognitive processes

Associations between eye movements and higher level perceptual and cognitive processes have been observed across different individuals. In particular, individual differences in memory among healthy younger and older adults (Chan et al., 2011; Firestone, Turk-Browne, & Ryan, 2007; Heisz & Ryan, 2011), older adults who show pathological aging (i.e. at-risk for mild cognitive impairment; Yeung, Ryan, Cowell, & Barense, 2013), and clinical groups such as amnesic individuals (Olsen et al., 2015, 2016; Ryan et al., 2000; Smith & Squire, 2017) have been associated with corresponding differences in eye movements.

For example, in the context of aging, eye tracking studies have shown that, when viewing faces, older adults tend to make fixations (Heisz & Ryan, 2011) and transitions between face features
(Firestone et al., 2007) at a higher rate than younger adults. Also, the aging literature contains evidence of older adults showing face recognition memory deficits that might relate to their reduced ability to bind face features – a process that would help form a memory representation of the face (Boutet & Faubert, 2006; Chalfonte & Johnson, 1996; Heisz & Ryan, 2011; Naveh-Benjamin Ben-Gurion et al., 2000). These deficits have often been attributed to age related atrophy of the hippocampal formation (Allen, Bruss, Brown, & Damasio, 2005; Pelletier et al., 2013; Walhovd et al., 2005). Therefore, corresponding age-related changes in eye movements are thought to be strategies that compensate for the reduced recognition memory deficits of older adults.

Similarly, altered gaze patterns seen in amnesic individuals imply an association between eye movements and memory processes while also helping researchers identify the behavioral manifestations of neural damage (Olsen et al., 2015, 2016; Ryan et al., 2000). For example, evidence from Olsen et al., 2016 showed disproportionately increased viewing of the eyes (during face viewing) and lack of an association between overall face viewing and subsequent recognition in an individual with developmental amnesia; suggesting such effects were related to the amnesic individual’s reduced hippocampal volume. Collectively, by presenting accounts of differential eye movement-memory associations across individuals, these studies highlight the ability of measures of eye movements (i.e. number of gaze fixations, transitions etc.) to index higher level perceptual and cognitive processes, and to outline the functions of underlying neural regions.

1.3 Static measures of eye movements

Various measures of eye movements have been used in the literature to depict the nature and time course of viewing in different kinds of experiments (e.g. free viewing vs performing an
experimental task). For example, viewing behavior has been summarized through eye movement measures including the number of gaze fixations made on stimuli (Chan et al., 2011; Heisz & Ryan, 2011; Henderson et al., 1999; Olsen et al., 2016; Parker, 1978; Yeung et al., 2013), number of transitions (saccades) made between stimulus features (Firestone et al., 2007; Loftus & Mackworth, 1978) and the extent/amplitude of saccades (Antes, 1974; Henderson et al., 1999). Other measures that have summarized different aspects of the time course of viewing, include average viewing time metrics (Antes, 1974; Buswell, 1935; Firestone et al., 2007; Mackworth & Morandi, 1967) and average gaze fixation duration measures (Antes, 1974; Loftus & Mackworth, 1978).

Definitions of selected static measures taken from Hannula et al., 2010 are presented below:

**Number of gaze fixations:** number of discrete pauses of the eyes for a display

**Number of saccades:** number of transitions made by the eyes between discrete regions

**Saccade amplitude:** distance traversed between successive fixations – in degrees per second

**Average fixation duration:** Mean length of time in which the eye pauses on a display. Fixation durations are usually 200-300msec long.

It is important to note that these measures provide a one-dimensional or static interpretation of the process of viewing. For example, the number of saccades within a given time duration will only provide a summary of the number of ballistic movements made by the eye when viewing a stimulus within the defined time frame. Therefore, these measures do not necessarily provide a record of real-time viewing behavior in the sense of capturing how visual information is continuously encoded by eye movements as they unfold over time.
These static measures have been directly associated with higher level perceptual and cognitive processes. For example, research shows that the overall number of gaze fixations made on stimuli is significantly associated with subsequent recognition performance in both older (Chan et al., 2011) and younger (Olsen et al., 2016) individuals. In these studies, the number of gaze fixations served to index recognition memory. However, it is important to note that these perceptual and cognitive processes are continuous and dynamic in nature (Aks, 2008; Spivey, 2007). Therefore, the static metrics of eye movements might not be fully capable of indexing the continuous nature of these processes. The dynamic nature of the perceptual and cognitive processes that have been associated with or indexed by eye movement measures have previously been studied extensively through Non-linear Dynamic Systems (NDS) methodologies.

1.4 Non-linear Dynamic System (NDS)

NDS is a general systems theory for describing, modeling, and predicting processes that are subjected to change over time and/or due to various influences (Gregson & Guastello, 2011). While linear models assume only one type of change and proportionality between inputs and outputs, NDS assumes that input and output relations need not be proportional and there are many possible patterns of change (Gregson & Guastello, 2011). Identifying these types of change contributes to a better understanding of the phenomenon at hand, provided that appropriate NDS analyses are used to identify the literal dynamics of the phenomenon (Gregson & Guastello, 2011). NDS analyses have been developed from concepts and/or influences derived from the fields of mathematics, biology, physics and social sciences (Gregson & Guastello, 2011). In the field of mathematics, it is possible to generate numbers from a model of one’s choice and analyze the series for properties of one’s choice (Gregson & Guastello, 2011). Therefore, there is no shortage of data in the series, because the numbers were made up and
noise can be deliberately put in to test a new procedure (Gregson & Guastello, 2011). Therefore, the simulation data that mathematicians deal with often fall outside the zone of real world data (Gregson & Guastello, 2011). However, in the science and social sciences, the data are derived from unknown models and consist of inherent noisy regions and complicated transient dynamic effects (e.g. temporary change in dynamics followed by reversion to the original dynamic pattern; Gregson & Guastello, 2011). Hence, numerous techniques have been developed to capture the NDS properties of data that are not constrained by specific models used in mathematics (Gregson & Guastello, 2011). These techniques include dimension analyses and information/entropy functions (Gregson & Guastello, 2011). Similarly, various techniques such as computational filtering have been used in NDS to deal with signal noise (Gregson & Guastello, 2011).

The NDS approach has led to a goal shift from simple cause effect relations captured by linear models to systems dynamics that consider time as a critical variable of interest (Aks, 2008; Ogmen, Bruno & Breitmeyer, 2006). In recent times, the importance of dynamics and dynamical recurrence in mental states is being increasingly acknowledged by researchers investigating human perceptual and cognitive behavior, resulting in a greater need to use NDS to probe these aspects of human behavior. Furthermore, the NDS approach provides an account of how unique behaviors emerge from a common mechanism under different conditions (Aks, 2008; Gregson & Guastello, 2011). This then could lead to a general theory of perceptual-cognitive behavior that accounts for and fits context specific behavior into a broader framework (Aks, 2008). Since NDS essentially provides a broader context/explanation/theory to specific and varying behaviors and explains how behaviors change across different conditions (e.g. with time), one could also use NDS techniques to predict future behavior (behavior you would expect at a future point in time and/or in a novel condition; Aks, 2008).
1.5 Characterizing the dynamic nature of perception and cognition using a NDS approach

Many studies have explored the continuous and dynamic nature of perceptual and cognitive processes (Spivey, 2007; Spivey & Dale, 2004, 2006) using NDS techniques. These studies have shown that perception is a dynamic entity that has a flow-like nature with reciprocal relationships with the brain, body and external environment (Valera, 1999; Valera, Thompson & Rosch, 1991). In other words, the behavioral patterns that are fundamental to perception are both continuous and relatively discrete in form, highlighting a given perceptual system’s natural variability over time (Aks, 2008). These behavioral patterns evolve over time, are dependent on changes in surrounding context as well as various internal and external conditions and consist of and reflect the patterns of information driving human perception (Aks, 2008).

Stimulus sensation is followed by perception, cognitive interpretation of the stimulus and motor action; thus, the progression of these processes can be treated as a continuous (not necessarily always sequential) process that arises from complex short and long range neuronal interactions (Aks, 2008; Spivey & Dale, 2004; Spivey, Richardson, & Dale, 2008). In other words, sensory impulses flow into perceptual regions of the brain, which also get feedback impulses from cognitive areas of the brain (Spivey & Dale, 2004). This blended (sensory and cognitive) pattern of activation spreads to other higher cognitive areas, and from there, to motor areas of the brain that facilitate motor action in response to the initial stimulus (Spivey & Dale, 2004). The motor output will also feedback and alter sensory information received by the brain – creating a circuit loop (Spivey & Dale, 2004). Based on this general view of impulse transmission in the brain, cognitive processing of information (sensory reception, leading to perception and cognitive interpretation) can be best described as a continuum as opposed to a process consisting of
discrete mental states (see Spivey & Dale, 2004 for simulations of the temporal continuity seen in cognitive processes including spoken word recognition, sentence processing, object/face recognition etc.), allowing us to explore cognitive processes from a dynamical systems point of view (Spivey & Dale, 2006).

In particular, real-time cognition can be described as a pattern of neuronal activity that continuously changes with time while traversing a state space (multi-dimensional space with distinct mental states; Spivey & Dale, 2006). The mental state most frequently visited by the mental trajectory at any given time constitutes the outcome of a given cognitive (mental) process. For example, when participants are showed pictures of a candle and candy and are asked to pick the candy, they initially look at the candle because it is phonologically similar to the candy, but eventually and more frequently select the candy (Spivey & Dale, 2006). In this case, the mental trajectory/activity was attracted to the candy, and recognition of this object is considered the final outcome/destination of this process/task (Spivey & Dale, 2006). However, the path taken to reach this destination still consists of intermediate regions of that space that reveal graded mixtures of mental states that need to be analyzed thoroughly in order to get a clear representation of the changes in cognitive processes that occur in real time (Spivey & Dale, 2006). It is important to note that the transitions between these mental states consist of continuous nonlinear change (Spivey & Dale, 2006).

Furthermore, in comparing results from NDS analyses with the information gained from conventional experiments and general linear model (GLM) analyses in cognitive research (Gilden, 2001; Sornette, 2004; Van Orden, Holden & Turvey, 2005), some claim that NDS might provide a better understanding of subtle behavioral patterns present in fluid systems (Aks, 2008). This is because conventional experiments and GLM models involve framing studies in a
particular way at the cost of bias (Aks, 2008). When our assumptions and the manner in which we frame the study drive analyses, certain intrinsic dynamics/patterns that may be signatures of fundamental perceptual and cognitive processes might be overlooked (Aks, 2008; Spivey & Dale, 2006; Gregson & Guastello, 2011). In contrast, NDS methods that do not include artificial constraints that can interrupt temporal effects, can be used to analyze the intrinsic/natural dynamics of processes in a more unbiased way (Aks, 2008). Given that higher-level perceptual and cognitive processes are best characterized as dynamic processes from a NDS standpoint, it is also important to understand and quantify measures that index these processes in the same manner. For example, it is important to analyze the continuous and dynamic nature, including spatial and temporal properties of eye movements in order to understand how they index perceptual and cognitive processes over time.

1.6 Characterizing the dynamic nature of eye movements using a NDS approach

The dynamics of eye movements have been analyzed from a NDS standpoint on the basis of several characteristics that make them suited for such an analysis (Aks, 2005, 2008; Aks, Zelinsky, & Sprott, 2002; Spivey, 2007; Spivey & Geng, 2001; Spivey et al., 2008; Aks, 2011; Fairbanks & Taylor, 2011). Eye movements have the capacity to serve as real-time behavioral records and are highly coordinated with other measures such as attention, reaction time and physiological parameters (Aks, 2008). They also serve as the interface between sensory and motor processes (Aks, 2008; Spivey & Geng, 2001). Due to these characteristics, the visual (oculomotor) system is influenced by many internal and external factors and displays critical properties of complex systems, including efficiency and coordination (Aks, 2008). Accordingly, the dynamic visual patterns that are produced by the visual system can be investigated through
an analysis of the range and pattern of correlation across a behavioral time series (e.g. during visual search) and unique temporal patterns that might be indicative of system function (Aks, 2008). These investigations aim to identify a “statistical form of memory” or meaning that can be derived from gaze signals, and can be done systematically through a NDS approach (Aks, 2008). Hence, in the field of visual perception, NDS analyses have been used to investigate crucial factors including of the nature of spatial and temporal patterns of eye movements (Fairbanks & Taylor, 2011), that are otherwise dismissed in GLM research (Aks, 2008).

1.6.1 Overview of the nature and dynamics of eye movements

NDS analyses suggest that eye movements produce recurrent gaze patterns that may be hidden amidst gaze signals (Wu, Anderson, Bischof, & Kingstone, 2014; Aks, 2011). These recurrent patterns (signal fluctuations) occur across many frequencies, or inversely, time intervals, i.e., gaze re-fixations occur in close and large temporal intervals from initial fixations (Wu et al., 2014), and their contributions to the overall viewing process follows a power law scaling relationship (Fairbanks & Taylor, 2011). In other words, the mean square amplitude of gaze signals varies as a power of frequency, and therefore can be described by specific power laws. These plots can be obtained from a Fast Fourier Transform (FFT) analysis applied on gaze signals. A linear function on the corresponding log-log plot of power (mean square amplitude) versus frequency will indicate the presence of a power law scaling relationship (Aks et al., 2002; Fairbanks & Taylor, 2011). Similar to other processes of nature (physiological or social), the power law scaling behavior that is shown by gaze signals does not feature a single characteristic timescale, and therefore, gaze signals display similar characteristics over many timescales, showing scale invariance (characteristic of “scale-free” systems; Fairbanks & Taylor, 2011). These temporal dependencies shown by gaze signals also involve changes in signal variability over time in subtle
and non-proportional ways due to successive behaviors influencing each other (Aks, 2008; Fairbanks & Taylor, 2011). In addition to the recurrent patterns that occur over many time scales, gaze signals also show repeated patterns in their spatial trace, which provides a fractal identity (spatial analogue to the power law behavior; Fairbanks & Taylor, 2011; Aks, 2008; Aks et al., 2002). A fractal identity or dimension can be defined as a ratio of complexity that explains how the detail of a fractal signal pattern changes with the scale at which it is measured (Fairbanks & Taylor, 2011). The fractal identity of gaze signals changes with the spatial context of viewing behavior, e.g. the change in the size of the external viewing space (Aks, 2008; Fairbanks & Taylor, 2011). For scale invariant processes such as viewing that show temporal and spatial dependencies, the repetition of signal structures across many scales (temporal and spatial) builds immense complexity in the resulting signal pattern by placing it in an intermediate balanced state between perfect order and complete disorder (Fairbanks & Taylor, 2011; Aks, 2008; Aks et al., 2002). This complexity is central to the behavior/process mediated by the system that produces the signal pattern (Fairbanks & Taylor, 2011).

1.7 Quantifying the temporal and spatial dynamics of eye movements

As mentioned earlier in this review, the presence of repeated patterns in gaze signals produced over the course of viewing process makes it a scale-invariant process. Hence, as with all scale-invariant processes, gaze signal patterns that arise from viewing show power law scaling behavior. The type of power law scaling behavior can be quantified by one or more power law exponents and will depend on the type of gaze signal patterns (e.g. absolute gaze position data versus displacement between gaze positions) and on the visual task at hand (free viewing vs visual search; Aks et al., 2002; Fairbanks & Taylor, 2011). Since gaze signal patterns are
essentially produced by eye movements, the power law behavior seen in gaze signal patterns actually reflects (serves as a marker of) the corresponding scaling behavior of eye movements (Fairbanks & Taylor, 2011). Therefore, this power law scaling behavior could explain how eye movements change across space and time, during viewing. The spatial and temporal changes in eye movements can be collectively called eye movement dynamics.

As mentioned at the beginning of this review, eye movements consist of long ballistic trajectories called saccades. However, they also consist of smaller motions called micro-saccades that occur during dwell periods (summation of duration across all gaze fixations when the gaze remains stationary on the stimulus; Fairbanks & Taylor, 2011) to ensure that the retina of the eye does not get desensitized. Micro-saccades typically occur over an angular range of 0.5 degrees (in a laboratory setting where subjects are shown stimuli on the screen, this value translates to ~15 pixels; Fairbanks & Taylor, 2011). Since saccades and micro-saccades are produced by different physiological mechanisms and serve different purposes in the visual system, their power-scaling behaviors will also be different (Fairbanks & Taylor, 2011). Generally, when analyzing the scaling behavior from spatial gaze patterns, a size (spatial) scale crossover of ~15 screen pixels can be seen between the saccadic and micro-saccadic processes of eye movements (Fairbanks & Taylor, 2011).

1.7.1 Effective dimensions of power laws that characterize the scaling behavior of eye movements

In order to understand relative differences in the saccadic and micro-saccadic scaling behavior of eye movements, Fairbanks & Taylor (2011) analyzed eye movements recorded during a visual search experiment. Specifically, the data presented and analyzed in this book chapter relate to the behavior of the eye as it searches for valuable information that is hidden or embedded in a clutter
of highly complex background information (Fairbanks & Taylor, 2011). It is important to note that physical processes like eye movements have inherent upper and lower cutoffs of magnification, making them “limited-range fractals” (Fairbanks & Taylor, 2011). In other words, one cannot increase the resolution of the eye movement data collected from an experiment or the length of the trace, for magnification purposes, without seeing a shift in physical processes (the expected physical process, e.g., saccadic eye movements will not necessarily remain the same after magnifying the data through a change in resolution, because invariably a different process would have taken over and will now be captured at the magnified region; Fairbanks & Taylor, 2011). Hence, in order to differentially analyze the temporal and spatial scaling properties of gaze traces, the $D_T$ and $D_S$ exponents calculated from the gradient ($\alpha$) of limited-range scaling plots (power spectra that only show the signal region that follows power laws, i.e. linear regions of a log-log plot of signal power versus frequency) are used as “effective dimensions” of power laws (Fairbanks & Taylor, 2011). Specifically, $D_T = (5 - \alpha)/2$ and $D_S = 2/(\alpha - 1)$. These exponents are specific to signal types (e.g. white signals, different colored signals etc., mentioned in detail later in this review). For instance, $D_T$ lies between 1 and 2 for timeseries of colored signals and has a value of 2 for white noise. Similarly, $D_S$ is 2 for white, pink and brown signals and lies between 1 and 2 for black signals. For our purposes, the temporal and spatial scaling properties of eye movements, documented by Fairbanks & Taylor (2011) and presented next in this review are also quantified and explained by the exponents: $D_T$ and $D_S$, respectively.

1.7.2 Temporal dynamics of eye movements

Fairbanks & Taylor (2011) have shown that individual micro-saccades operate over a time period (time between successive micro-saccades) of ~10-20 ms during dwell periods (periods of relative motionlessness) of approximately 0.4s. On the other hand, saccades are of more interest
to vision researchers, as eye movement patterns created by these larger motions dictate the task at hand (e.g. visual search) and operate over longer timescales (time between successive saccades is longer than that between successive micro-saccades). Hence, these longer timescales of interest dominate the temporal trace (e.g. plot of the x coordinate of eye position versus time) of eye movements and are formed by an interplay between dwell periods and saccades (Fairbanks & Taylor, 2011). Accordingly, in calculating the $D_T$ parameter for the temporal trace recorded in their study, Fairbanks & Taylor (2011) found this parameter to be 1.2. This value indicates that the temporal scaling behavior of eye movements that they recorded can be characterized by ‘black power laws’. Note that power law scaling behavior of signals are differentiated based on the relative number of signal structures that correlate with one another over longer time intervals, or in other words, based on the degree of signal power at lower frequencies (Fairbanks & Taylor, 2011). As such, colored power laws (e.g. pink, brown, black) have higher signal power at lower frequencies whereas white noise has no correlation between signal structures over any time interval, and as a result, the signal power remains constant across all frequencies (Costa et al., 2005; Fairbanks & Taylor, 2011). These colored power laws distinguish the informativeness of signals, e.g. white noise is considered non-informative compared to pink (Costa, Goldberger, & Peng, 2005). This then provides reason/basis to quantify the degree of informativeness of these colored signals using different measures.

Fairbanks & Taylor (2011) also refer to the work done by Aks et al. (2002) in their book chapter, where observers were asked to search for embedded icons. Interestingly, results from this study indicated that the gaze fixations showed Brown power law behavior (Fairbanks & Taylor, 2011; Aks et al., 2002). Moreover, when Aks et al. (2002) analyzed the power law behavior exhibited by the gaze displacement data (distance between successive fixations or relative gaze position data), they observed that data at lower frequencies predominantly followed pink laws (highly
correlated structures occurring over longer time intervals). They further observed that the higher frequencies showed brown color law behavior indicating the presence of highly correlated structures occurring over shorter time intervals. Collectively, these results suggest that eye movements can be defined by colored power laws featuring highly correlated (similar/repeated) signal structures or events present over shorter (e.g. brownian/brown law) and relatively longer (e.g. pink law) time intervals, as opposed to white noise, which has no correlated structures at most time intervals. Overall, it seems like the exact type of colored power law would depend on the specific visual task at hand (e.g. visual search), the type of eye movement data (e.g. gaze fixations vs displacement between gaze fixations) being analyzed and the time scales (or frequencies) considered for analysis.

1.7.3 Spatial dynamics of eye movements

From the spatial eye movement trace (e.g. x coordinates of eye position vs y coordinates of eye position) obtained from their study, Fairbanks & Taylor (2011) observed that at around 0.5 degrees (18 pixels), a motion dominated by saccades crossed over to micro-saccades. This is quite close to the expected crossover size scale of ~15 pixels. In calculating the effective spatial dimension ($D_S$) at larger time scales dominated by saccades, they found this parameter to be 1.4, which corresponds to colored, specifically, black power laws ($2 > D_S \geq 1$; Fairbanks & Taylor, 2011; significance of this value is explained below). This value was consistent across subjects and insensitive to the $D_S$ value of the fractal pattern of the image being observed. A similar $D_S$ value ($D_S = 1.5$) was obtained when subjects viewed different types of fractal images (e.g. observing fractal images separated by a 30sec checkerboard pattern) or fractals of different colors ($D_S = 1.4$). In analyzing the reason why the human eye follows a fractal trajectory with an inherent $D_S$ value of ~1.5 in all conditions, Fairbanks & Taylor suggested that, based on the
mathematics and geometry of fractals, this value allows the eye to cover the visual terrain efficiently, through ‘enhanced diffusion’ shown by eye movements, when searching for information. Enhanced diffusion is an optimal search strategy compared to forming trajectories that are confined to a small space (Brownian motion with a $D_s = 2$), or a quick but simple straight trajectory that does not involve exploration of the 2D plane ($D_s = 1$). This interpretation is also confirmed by previous research that examined whether there was a marked change in physiological responses of people viewing mid-$D_s$ fractals (i.e. viewing fractal images whose spatial scaling properties matched those of the human eye; Hagerhall et al., 2008; Taylor, 2006). Results from this research showed a reduction in the physiological stress levels of observers when viewing such images (Hagerhall et al., 2008; Taylor, 2006), confirming the adaptive role of $D_s = 1.5$ in active viewing. These studies also emphasize the importance of understanding and applying the scaling properties of eye movements in vision research (Fairbanks & Taylor, 2011).

Similarly, results from two analyses conducted by Aks et al. (2002) that explored the spatial pattern of eye movements also show that eye movements might have a fractal identity that is equivalent to colored power laws. First, in plotting the x eye position data (x coordinate values of gaze fixations) against y eye position data (y coordinate values of gaze fixations), Aks et al. (2002) observed that differences across y positions increased over time, whereas differences across x eye positions decreased over time. Therefore, they suggested that such changes in means and variance of gaze fixations with fixation duration is characteristic of fractal structures.

Second, they used the Iterative Function Systems (IFS) test to identify the nature of correlation seen in eye position data. IFS provides sensitive tests that show deviations from randomness in data (see Aks et al., 2002 for a detailed explanation on the technique). The IFS procedure when applied on data, produces clumped patterns if the data consists of colored noise and homogeneously filled spaces when the data is uncorrelated, i.e. white noise (Aks et al., 2002).
Specifically, within the spectrum of colored noise, brown noise is usually characterized by clumped patterns of data along the diagonals or sides of the IFS graphical output (Aks et al., 2002). Pink noise is depicted by self-similar repeated triangular structures of different sizes that accumulate near the diagonals (Aks et al., 2002). Meanwhile, white noise uniformly fills the space of representation (Aks et al., 2002). Accordingly, when IFS was applied on x and y position data independently, Aks et al. (2002) noticed evidence of a short-term and highly correlated pattern of brown noise (pre-dominant) as well as weaker long-term correlations associated with pink noise. Similarly, when IFS was applied on data corresponding to the net displacement between eye positions, they observed a clear and diffused colored noise pattern. Collectively, these results suggest that similar to the temporal scaling behavior of eye movement signals, the overall spatial scaling behavior has the fractal identity of colored power laws, in contrast to white noise. Again, as in the case of temporal dynamics, the type of colored fractal laws followed by eye movements might depend on the type of eye movements investigated (e.g. absolute versus relative gaze position data) and the visual task (e.g. visual search).

1.8 Importance of quantifying spatial and temporal dynamic properties of eye movements

In general, studies described above suggest that correlated (repeated/similar) structures, the pattern of which can be characterized by one or more colored power laws, are present in gaze signals. These correlations can be identified across many different spatial (e.g. signal amplitude differences) and temporal (longer versus shorter time intervals) scales. The nature of these correlations seems to be dependent on the type of gaze signals explored, spatial and temporal scales being considered in the investigation and the visual task at hand. These repeated structures provide structural complexity to the gaze patterns (Fairbanks & Taylor, 2011; Aks, 2008; Aks et
al., 2002). The presence of highly correlated signal structures in gaze data across many scales (spatial and temporal) suggest that memory exists in gaze signals (Aks et al., 2002). Furthermore, the fact that the contingencies across fixations differ by the type of gaze signals, spatial and temporal scales analyzed, and visual task at hand indicate that the memory that exists in gaze signals varies with these parameters (Aks et al., 2002). Collectively, these observations make viewing an informative and meaningful process to investigate (Fairbanks & Taylor, 2011). In particular, quantifying the structural complexity of signal structures contained in gaze patterns produced by eye movements over time should help explain the dynamics of eye movements. Since eye movements facilitate viewing, this approach will contribute to an understanding of how the process of viewing unfolds over time. Moreover, in relation to the association seen between eye movements and higher-level processes (cognition, perception), a measure of eye movements that quantifies the structural complexity and memory that exists in gaze patterns as they unfold over time can better index the manner in which perceptual and cognitive processes influence and are informed by eye movements over time. One way of quantifying the structural complexity of physiologic or physical signals is by measuring the degree of signal regularity, i.e. entropy. Accordingly, in the current thesis, we have treated gaze patterns as signals and applied an entropy measure called Multi-Scale Entropy (MSE) to quantify the structural complexity arising from repeated/correlated visual events (e.g. saccades) present in gaze signal patterns, as these patterns unfold over time. We expect this measure to serve as a measure of eye movement complexity that would help quantify viewing dynamics, i.e. viewing changes that occur over time.
1.9 Relationship between signal regularity, entropy and information

Entropy based measures are traditionally known to quantify the degree of regularity, order (Guastello, 2011) or predictability of signals. From a NDS point of view, entropy measures evaluate the appearance of repetitive patterns present in signals, including random and/or deterministic structures. These structures present in signals are useful in quantifying the complexity of the system that produces the signals (Costa et al., 2005; Aks, 2011). Higher entropy is considered to be indicative of greater irregularity or disorder, which could lead to greater variability and behavioral flexibility over time (Aks, 2011; Guastello, 2011; Costa et al., 2005).

Entropy has also been associated with the concept of information since the origin of a problem connected to the engineering of long-distance telephone systems in 1948 (Shannon, 1948; Guastello, 2011). In Shannon’s information theory he refers to a bit of information as the amount of information that is required to determine a dichotomous outcome (Shannon, 1948; Guastello, 2011). Therefore, for an outcome involving N equiprobable states, N-1 bits of information are needed (Shannon, 1948; Guastello, 2011). For a truly random system that produces a long string of observations, the number of bits of information that becomes necessary to determine the series of observations would be approximately equal to the number of original data points in that series (Shannon, 1948; Guastello, 2011). More specifically, Shannon’s entropy (H$_S$), which quantifies the pattern (regularity) of datapoints present in a series has been usually directly equated to the information required to describe those patterns (I; Shannon, 1948; Guastello, 2011). Therefore, in relating the concepts of entropy, regularity of a timeseries (based on symbolic dynamics) and
information, the regularity of a timeseries as quantified by entropy measures, is also a measure of information contained in the timeseries (Guastello, 2011; Sarkar & Barat, 2008).

1.10 Usage of Shannon’s entropy and its derivatives

The applications of Shannon’s entropy in their simplest form have involved a single set of discrete states and probabilities that are associated with them, where the states tend to appear in a one-dimensional time series (Guastello, 2011). However, since its origin, the algorithm has been modified and concept expanded in ways that covered continuous variables, more than one dimension and real time physiological and physical timeseries (Guastello, 2011). The algorithms derived from Shannon’s entropy have been used to address different research analytical needs depending on their distinct features. Based on relevance to the current review, comparisons have been made between a few selected types of entropy-based algorithms that are derived from Shannon’s entropy.

1.10.1 Topological entropy (Kolmogorov-Sinai entropy)

If we were to consider low dimensional chaotic systems, which generate random sequences from simple functions, a small amount of information can explain a lot of activity (Guastello, 2011). These systems therefore produce information when in motion (Guastello, 2011). In such systems information and entropy become the same entity that is inherent to the system, which is essentially what Shannon proposed. This concept was also captured by an algorithm devised by Prigogine, called topological entropy, also called Kolmogorov-Sinai (K-S) entropy (Guastello, 2011; Nicolis & Prigogine, 1989). However, Prigogine in his K-S entropy algorithm extended the idea behind Shannon’s entropy to a multidimensional array with continuous data that is represented as discrete states (Guastello, 2011). Although this algorithm has been useful in
classifying low dimensional chaotic systems, it cannot be applied on experimental data because it gives a result of infinity for a process that contains superimposed white noise of any magnitude (Costa et al., 2005; Guastello, 2011). Therefore, approximate entropy ($A_E$) measures were derived from Shannon’s entropy to account for white noise that is bound to be present in real world signals.

### 1.10.2 Approximate entropy measures and sample entropy

In order to quantify the entropy of real world experimental data consisting of short and noisy timeseries, Pincus (1991) devised a family of measures called approximate entropy ($A_E$; Costa et al., 2005; Guastello, 2011) that is not susceptible to the influence of noise. $A_E$ was created to separate order from disorder, where low $A_E$ values correspond to order resulting from deterministic (ordered) processes, and higher $A_E$ values marks disorder/irregularity (Pincus, 1991; Guastello, 2011; Costa et al., 2005). This algorithm suggests that points in an ordered system would initiate close together and still sustain the closeness as the time series evolves (Guastello, 2011). A modification to the $A_E$ algorithm was later proposed by Richman & Moorman (2000), leading to the formulation of sample entropy ($S_E$). The advantages of $S_E$ over $A_E$ include its ability to be less dependent on the length of the timeseries under investigation and showing consistency over a wide range of parameters that are involved in entropy calculations (Costa et al., 2005; Richman & Moorman, 2000). These parameters include: the number of datapoints that make up the timeseries, number of adjacent points that are compared for similarity over the course of the timeseries, and the range within which the absolute magnitude of sets of adjacent points (e.g. magnitude of signal amplitude) are considered to be similar (Guastello, 2011; Costa et al., 2005).
Although A_E and S_E algorithms enable the computation of entropy from real life timeseries by being immune to the inherent noise (which was a concern with previous entropy-based measures such as K-S entropy), they are, by definition, measures that quantify the degree of randomness (or inversely, regularity/orderliness; Costa et al., 2005). However, regularity and structural complexity do not have a straightforward correspondence between them (Costa et al., 2005). Complexity is associated with “meaningful structural richness”, which in contrast with random signals, displays relatively high regularity (Costa et al., 2005). Complex signals often contain structures that correlate at multiple spatiotemporal scales and are also referred to as colored noise (Costa et al., 2005). Entropy based metrics such as S_E and A_E are not capable of distinguishing complex timeseries from random timeseries that follow white power laws (Costa et al., 2005) and can easily assign higher entropy values to randomized timeseries, including less adaptive/complex pathological timeseries, relative to complex timeseries. As mentioned in Costa et al., 2005, a reason why this is the case is that these metrics are based on a single scale and are not capable of accounting for features/correlation between structures present in a complex physiological or physical timeseries on scales other than the shortest one (Costa et al., 2005).

Since it is generally accepted that perfectly ordered and completely disordered timeseries do not contain complex structures that carry meaning, it is important to derive a metric that assign high values to colored noise and assign low values to completely ordered (regular) or disordered signals (white noise; Costa et al., 2005). Accordingly, for physical systems, a complexity measure devised by Zhang (1991) appeared to address this issue.

1.10.3 Zhang’s entropy

Zhang’s complexity measure involves computing the sum of scale-dependent entropies, which takes into account the information contained in multiple scales (Costa et al., 2005; Zhang, 1991).
This then allows the two extreme limits (complete order and disorder) to vanish. However, since Zhang’s method is based on Shannon’s definition of entropy, it is only capable of matching the physical data (timeseries) to a discrete symbolic sequence with sufficient accuracy when the physical data is relatively noise-free. This then limits its use on real world physiological signals, which as described previously in the context of K-S entropy tend to be quite variable (including noisy regions) and have finite length.

Accordingly, Costa et al., 2005, addressed the need to create a metric that could be applied on highly variable real world physiological and physical signals, and has the capacity to quantify their complexity by assessing entropy over multiple scales. To this end, they created the Multi-Scale Entropy (MSE) measure based on both Pincus’s and Zhang’s measures (Costa et al., 2005).

1.11 Multi-Scale Entropy (MSE)

MSE is a signal complexity metric that distinctly quantifies the complexity, or the degree of regularity of patterns of signal structures present in physiological or physical signals of finite length across multiple timescales (Costa, Goldberger, & Peng, 2002; Costa et al., 2005; Escudero, Abásolo, Hornero, Espino, & López, 2006; Heisz, Vakorin, Ross, Levine, & McIntosh, 2013; McIntosh et al., 2010, 2013).

MSE involves the computation of sample entropy (S_E) from multiple coarse-grained timeseries that correspond to specific scale factors (Costa et al., 2005). Specifically, S_E is first calculated on the original timeseries, which corresponds to a scale factor (timescale or τ) of 1 (Costa et al., 2005). This original timeseries is then divided into non-overlapping windows of length τ, and datapoints within these windows averaged, to produce coarse-grained timeseries that correspond to a given τ (Costa et al., 2005). The length of each coarse-grained timeseries would then be the
length of the original timeseries divided by the scale factor, $\tau$ (Costa et al., 2005). Therefore, MSE curves indicate the $S_E$ calculated from each of these coarse-grained timeseries (including the original timeseries that corresponds to $\tau=1$), as a function of $\tau$ (Costa et al., 2005). These curves are often used to analyze the relative complexity (entropy) of normalized timeseries (timeseries are normalized by keeping the variance of timeseries at all timescales equal to that of $\tau=1$, so that $S_E$ does not depend on absolute value of datapoints but rather on the sequential ordering or structural organization; see details about keeping parameter $r$ constant as a means of normalizing timeseries, in Costa et al., 2005) corresponding to specific timescales as opposed to being used to interpret the single complexity value assigned for a particular timeseries (Costa et al., 2005). Moreover, these timeseries comparisons are based on two guidelines. First, if the majority of scales ($\tau$ values) display a higher entropy for one timeseries than another, then the former is considered to be more complex than the latter (Costa et al., 2005). Second, given that entropy is a measure of information, a decrease in entropy values as a function of scales ($\tau$) would mean that the highest amount of information is contained in the smallest scale only (Costa et al., 2005).

Similar to Zhang’s entropy measure, MSE captures the sample entropy of timeseries over many scales and therefore can differentiate complex signals from periodic signals or white noise. Costa et al (2005) tested this prediction by applying MSE on simulated colored noise of the type $1/f$ versus white noise. Their results showed that although a higher entropy value was assigned for white noise compared to $1/f$ noise at scale 1, the entropy for the coarse-grained white noise monotonically decreased thereafter. However, the entropy of $1/f$ remained consistently higher than white noise at most coarser scales (Costa et al., 2005). Costa and colleagues (2005) suggested that the decrease in entropy of white noise at longer scales (higher $\tau$ values) could be due to a lack of new structures at larger timescales (as the $\tau$ or length of windows used to average
Data points increases, i.e. resolution decreases), resulting in the average value inside these windows converging to a fixed value. In other words, MSE, by assigning lower entropy values to larger timescales, in white noise, essentially highlights the lack of structures that correlate with each other at these scales. With respect to periodic (perfectly ordered) signals, the entropy will always be zero, since the occurrence of each datapoint is predictable (probability of occurrence is 1; Costa et al., 2005). Therefore, the entropy of periodic signals will never be larger than that of random signals (Costa et al., 2005). Due to the ability of MSE to assign higher entropy to structurally complex colored noise of the type $1/f$ and differentiate these signals from both white noise and periodic signals, it has been used in various research modalities to assess the complexity of different types of physiological and physical timeseries. Hence, in its various applications, MSE has been used to index the meaningful information contained in complex signals by measuring the entropy of these signals (Costa et al., 2005).

### 1.12 Previous applications of MSE

MSE has been applied on various physiological and physical signals to quantify the complexity of information contained in these signals. Since the health status of a system can be reflected by the complexity (dynamics) of signals produced by the system, MSE has been able to distinguish between normal and pathological states of systems through measuring the complexity of corresponding signals (Costa et al., 2005). Similarly, MSE has also increased our understanding of the how information is differentially processed by systems and carried in the signals produced by these systems (Heisz et al., 2013). Applications of MSE have also helped us understand how signal complexity and information processing changes with demographic factors such as age (McIntosh et al., 2010, 2013). Selected applications of MSE are presented below.
1.12.1 Application on cardiac inter-beat interval timeseries

MSE was applied on cardiac inter-beat interval timeseries derived from continuous electrocardiographic Holter monitor recordings of healthy subjects as well as subjects who showed cardiac dysfunction (congestive heart failure and atrial fibrillation; Costa et al., 2005). As expected, the signal entropy of healthy subjects showed an initial increase and stabilized at larger timescales, while the entropy quantified in congestive heart failure patients showed an initial decrease followed by an increase, but still remained below the MSE curve of healthy subjects at longer timescales (Costa et al., 2005). In contrast, MSE in patients who showed atrial fibrillation, which is considered to be a major cardiac arrhythmia, monotonically decreased with timescales, and produced a profile that was suggestive of white noise (Costa et al., 2005). Hence, these results showed that healthy heartbeat interval timeseries is more complex (as measured by higher entropy at larger timescales) than pathological timeseries (Costa et al., 2005). This result speaks to the ability of MSE to distinguish between healthy and diseased states based on signal complexity. It is important to note that the entropy at scale 1 was higher in subjects with cardiac arrhythmia compared to healthy controls, giving the false impression that the signal complexity of cardiac arrhythmia patients might be higher than healthy individuals. Therefore, this observation reinforces the importance of using MSE, which uses $S_E$ (equivalent to entropy value at scale 1) but recalculates it by scale – and this appears to be a better method to distinguish truly healthy versus pathological signal dynamics than only looking at $S_E$ at scale 1 (Costa et al., 2005).

1.12.2 Application on coding and non-coding sequences of DNA

Costa et al. (2005) applied MSE on DNA sequences that code for proteins as well as on non-coding sequences of DNA (junk DNA) that have recently been thought to have important uses in
gene transcription, DNA replication etc. although they do not directly code for proteins. DNA sequences were mapped to a numerical sequence based on the properties of DNA bases – Purine bases A and G were mapped to number 1 and Pyrimidine bases C and T were mapped to number -1. MSE was then applied on the resulting numerical sequence that served as the timeseries. Their results showed that junk DNA displayed higher entropy at larger timescales compared to DNA sequences that code for proteins, supporting several views of junk DNA containing important biological information and complexity that relates to mechanisms that control the integration of gene activity (Costa et al., 2005).

1.12.3 Application on neural signals

In neuroscience, MSE has been used to measure the complexity of neural signals such as those captured using electroencephalography (EEG) and magnetoencephalography (MEG). This has been done to identify differences in neural dynamics and cognitive information processing across different brain regions and task conditions. For example, Hesiz et al (2013) applied MSE on MEG data obtained while subjects recollected personal autobiographical/episodic (PE) and general semantic (GS) memories based on previous events that they themselves had recorded prior to the experiment (Heisz et al., 2013). Participants were also asked to rate the vividness of memories based on the overall richness of reexperiencing them (Heisz et al., 2013). As expected, these vividness scores confirmed that PE memories were more vividly reexperienced/recollected compared to GS memories (Heisz et al., 2013). In this study, MSE quantified the information processing capacity of brain regions that produced these neural signals (Heisz et al., 2013). In comparing MSE for the two conditions (PE vs GS), results showed that MSE was greater for PE than for GS and remained stable across multiple timescales in specific brain regions including the left dorsolateral prefrontal cortex and ventrolateral premotor cortex (Heisz et al., 2013),
leading to the conclusion that PE memories elicit a rich cognitive experience and a more complex brain response than GS memories (Heisz et al., 2013).

MSE has also been applied on MEG and EEG signal timeseries to identify the spatiotemporal dependencies, along with regional and task specificity of neural dynamics during healthy maturation (McIntosh et al., 2010, 2013; Misic, Mills, Taylor, & McIntosh, 2010). In particular, the research done by McIntosh et al. (2010) involved the application of MSE on EEG signals collected from children (between 8-15 years) and adults. Results from this study revealed that adults had higher neural MSE compared to children, indicating that brain signal complexity increases with maturation. In the same study, MSE was applied on EEG signals obtained from infants (in comparison to higher age groups, including adults) upon exposure to auditory and visual stimuli (McIntosh et al., 2010). While results from this study also showed a maturity related increase in MSE, they also further showed a greater complexity for visual stimuli, relative to auditory stimuli in infants (2-8 months, i.e. critical period for visual system development). This result indicated that maturity dependent increase in complexity might change across sensory domains (McIntosh et al., 2010). Finally, MSE was also positively associated with stable behavior, indicating that neural complexity might indeed be adaptive to some extent. In extending maturity related changes in neural complexity into the realm of healthy aging, past research showed that these changes become increasingly dependent on timescale (time interval being considered) as we age (McIntosh et al., 2013). Specifically, aging has been associated with higher entropy (complexity) at finer timescales and lower entropy at coarser timescales (McIntosh et al., 2013), suggesting that neural changes that occur in aging might result in more spatiotemporal dependencies and increased local processing (based on high entropy indexed at finer vs coarser timescales in older individuals; McIntosh et al., 2010, 2013).
Essentially, the above applications of MSE indicate that this metric can distinguish the complexity arising from correlated structures and corresponding information contained in signals that follow colored power laws, in particular, the form $1/f$, from white noise or periodic behavior. Since complex signals are often produced by healthy systems, MSE has the capacity to distinguish between healthy versus pathological systems.

1.13 Current attempt: application of MSE as an eye movement complexity measure

Provided that a given type (e.g. absolute position versus gaze displacement) of gaze signal patterns produced from eye movements follows colored scaling behavior as opposed to white noise, one could expect MSE to quantify the complexity (entropy) and meaningful information contained in gaze signals as they unfold over time. Since the current work is inspired by the previous usage of MSE on colored signals of the specific type $1/f$ (e.g. Costa et al., 2005) and contributes to previous research on quantifying $1/f$ type gaze signals (e.g. Aks, 2005), we chose to first identify if our gaze signal also belonged to this category of signals. Research done by Aks et al. (2002) showed that relative eye position (gaze displacement) signals recorded during a visual search task predominantly followed pink power law ($1/f$) behavior at low frequency regions (corresponding to longer time scales), and brown power law ($1/f^2$) behavior at higher frequencies. However, in the same study, they observed that absolute gaze position signals recorded from the same task predominantly showed brownian motion ($1/f^2$). Taken together, both results suggest that the type of color laws followed by gaze patterns might differ with the type of gaze signals being analyzed. Furthermore, the type of colored law behavior shown by gaze signals might also alter with the type of task. For example, Fairbanks & Taylor (2011) recorded gaze position signals from a visual search task involving searching for valuable information that
is hidden or embedded in a clutter of highly complex background information. Their temporal visual trace showed that these signals followed black power laws. This result contrasted with the brownian behavior shown by gaze position data recorded by Aks et al (2002), a study which involved searching for icons embedded among distractors. Since $1/f$ type signals indicate the presence of highly correlated structures at longer time intervals, they provide evidence for memory across these structures. This makes these signals meaningful and provide basis for applying MSE to quantify the structural complexity and information contained in the gaze signal patterns. Therefore, although MSE can be applied on any type of gaze signal, based on previous work, the information it quantifies is expected to carry meaning if derived from signals of the type $1/f$.

MSE, if successfully applied on gaze signal patterns, could serve to quantify the information encoded by eye movements over the course of viewing. This could then serve as an eye movement measure that contributes to an understanding of how eye movements guide viewing in a continuous manner. Finally, as we know, eye movements are guided by bottom up and top-down cues present in stimuli. Therefore, based on the bottom-up and top-down information contained in different stimuli, eye movements are expected to differentially encode information from stimuli during viewing. We could also then expect this information to be differentially quantified by MSE as a function of stimulus type.

1.14 Overview of thesis

The purpose of the current thesis is to apply MSE on gaze displacement signals (change in gaze position) produced from eye movements, in order to contribute to an understanding of how eye movements encode information as viewing patterns unfold over time. In the current study, gaze displacement signals have been computed from eye movements that were recorded while
participants viewed scenes that belonged to different categories. In order to justify using MSE to quantify the complexity of gaze displacement signals and meaningful information contained therein, we will first assess whether these signals follow pink (1/f) power laws. Once we confirm the power law behavior exhibited by our gaze signals, we will assess how sample entropy (information encoded by eye movements and contained in gaze signals) changes with gaze signals captured over different time intervals (finer versus coarser timescales). These time intervals reflect the different durations between important visual events, such as saccades. In other words, this analysis will help identify how information is differentially encoded by saccades that occur closer (finer timescales) versus further apart (coarser timescales) in time.

If MSE successfully quantifies the complexity of gaze signals by assigning high sample entropy to gaze signals captured over coarser timescales, we will then assess how MSE differentiates between gaze signals that are elicited by different types of scenes (e.g. fractals, scene depicting social interactions etc.). Specifically, the scenes used in the current study were obtained from a publicly available database of naturalistic and artificial scenes (Borji & Itti, 2015). These scene categories vary in informational content (e.g., people, places, and objects), and vary on a Normalized Scanpath Saliency (NSS) score (Borji & Itti; 2015). The various scene categories can give us a sense of the top-down relevance of scenes, whereas the NSS score is an index of bottom up information. Specifically, in Borji & Itti (2015), saliency models were used to obtain average NSS scores per image category based on bottom-up image features. Based on the availability of NSS scores per image category in http://saliency.mit.edu/results_cat2000.html (website associated with the paper), the models that we used to calculate average NSS scores (average NSS across models) per image category, for our study, are ITTI (Itti, Koch, & Niebur, 1998; model detects target from surrounding distractors based on unique target size, intensity, color and orientation), GBVS (Harel, Koch, & Perona, 2006; model uses centre to surround
preference, i.e. centre bias) and AWS (Garcia-Diaz, Leboran, Fdez-Vidal, & Pardo, 2012; model uses an optical variability estimate involving intensity, spectral wavelength and spatial frequencies). Also, it is mentioned in Borji et al. (2015) that NSS does not account for top down cues and NSS of all these models are similar to one another. By observing how MSE tracks with the different scene categories, we aim to find out how MSE varies with the top-down relevance of scenes. The average NSS scores of scene categories can be compared to the MSE of scene categories to understand how both measures compare with one another (i.e. whether and how MSE relates to bottom up content of images). Also, we can find out whether scene differences in MSE are reliable at a few selected versus all timescales.

Finally, in order to identify whether MSE might provide additional information to what can be derived from static eye movement measures (e.g. number of saccades, average saccade amplitude etc.), we will probe for differences in MSE across scenes that might remain after regressing out static measures from the gaze signals.
Chapter 2

Research aims and hypotheses

In this thesis we have contributed to cognitive and vision research by developing a novel metric to quantify the spatial and temporal changes (dynamics) present in eye movement data collected from empirical studies. This metric is analogous to metrics developed for neuroimaging studies. Current studies that use eye tracking to indirectly assess perceptual and cognitive processes typically explore their data from a “static” point of view (e.g. number of fixations/saccades, saccade amplitude, viewing time duration, average fixation duration). These measures, while serving to provide a summary regarding the manner by which viewing behavior unfolds, may not necessarily capture the inherent dynamics of gaze signals produced by eye movements. Hence, in an attempt to address this issue, we have applied a signal complexity metric called Multi-Scale Entropy (MSE) on eye movement signals (Costa, Goldberger, & Peng, 2002, 2005). We have chosen MSE as the ideal metric of eye movement signal complexity based on previous applications of MSE on other types of physiological signals that show comparable signal dynamics (Costa et al., 2005; Heisz, Vakorin, Ross, Levine, & McIntosh, 2013; McIntosh et al., 2013). These previous applications have highlighted the unique capabilities of MSE to quantify the dynamics of these signals.

Previous research studies have used non-linear dynamic system (NDS) methodologies to explore the nature of gaze signals and have documented key features that characterize the spatial and temporal dynamics of these signals. Viewers may exhibit recurrent gaze patterns; that is gaze fixations may re-sample previously viewed regions on a given image. The repetition of gaze fixations in their spatial trace (e.g. plot of x versus y coordinates of gaze position) gives eye
movement patterns a fractal identity (Aks, 2008; Aks, Zelinsky, & Sprott, 2002; Fairbanks & Taylor, 2011). Signals are considered to be fractal if the structural detail changes with the scale at which it is being analyzed and fractal indices can quantify how this detail changes with scale (e.g. spatial resolution of signal). Such re-sampling behavior may occur at close and/or distant time intervals from the initial sampling (Wu, Anderson, Bischof, & Kingstone, 2014; Aks, 2011), making viewing a scale-invariant process that shows similar characteristics over multiple timescales. Moreover, these fixations can influence each other in non-proportional ways (Fairbanks & Taylor, 2011). These types of signals follow a power law scaling behavior involving a change in the mean square signal amplitude as a power of signal frequency (Fairbanks & Taylor, 2011). Based on the above evidence, gaze signals could be expected to show similar patterns over multiple time scales, and therefore may not be best characterized by obtaining summary metrics over one particular timescale (e.g., the full trial).

Accordingly, the repetition of gaze signal structures over spatial and temporal scales builds complexity in the resulting gaze patterns, placing them in an intermediate state between complete order and disorder (Aks, 2008; Aks et al., 2002; Fairbanks & Taylor, 2011). The exact type of complexity, also referred to as signal color seems to vary with the visual task at hand (e.g. visual search task), scales being analyzed (e.g. frequency range) and the type of signals (e.g. absolute gaze position, distance between gaze position etc.; Aks, 2005; Aks et al., 2002; Fairbanks & Taylor, 2011). A particular type of signal color that has gained attention in the NDS-vision literature and relevant to the current study is pink ($1/f$) noise. This type of signal is characterized by structures that correlate with each other over longer time intervals (Aks et al., 2002; Costa et al., 2005). In particular, work done by Aks et al. (2002) showed that displacement between gaze positions recorded from a visual search task appeared to predominantly be of the form $1/f$ at low frequency regions, indicating that highly correlated structures were present over longer time
intervals. Hence, they claimed that this finding indicates that these contingencies across fixations, present over long time intervals could indicate that a form of spatial memory might exist across fixations, making gaze signals of the type 1/f meaningful.

Costa and colleagues (2002, 2005), who first devised the MSE algorithm, validated its usefulness as a signal complexity metric by contrasting the manner in which it assesses the complexity of 1/f signals and white noise. White noise is considered to be meaningless in comparison to 1/f signals because they are purely random and might only contain signal components that correlate with each other over shorter time periods (short-term local correlations). MSE, in theory, assesses the complexity of signals, by quantifying the sample entropy of a given signal time series, at multiple timescales (Costa et al., 2005). Entropy, which quantifies the degree of regularity of signals has been equated with concepts of information (Shannon, 1948). This has led to the idea that the regularity of a given time series, quantified by entropy, relates to the amount of information required to describe it (Guastello, 2011; Sarkar & Barat, 2008). Hence, MSE, which is essentially a metric of entropy, captures the information contained in real world physiological or physical signals of finite length, or in other words, the information produced by a given system and carried in its signals (Costa et al., 2002, 2005). Costa and colleagues (2005) showed that the uniqueness of MSE (in comparison to other entropy measures) lies in its ability to differentiate the useful information contained in complex 1/f signals from the relatively low degree of signal complexity and useful information contained in periodic (regular) signals and white noise (Costa et al., 2002, 2005). MSE highlights this difference by assigning higher entropy values to 1/f signals at longer time scales (signal captured at a lower frequency/resolution) relative to periodic signals or white noise. Accordingly, based on these findings, we hypothesized that if gaze displacement signals derived from a particular
visual task can be characterized as 1/f type signals, then MSE can be used to quantify the complexity and meaningful information contained in these signals.

In the current study, we applied MSE on eye movement data collected from individuals while they freely viewed different types of scenes obtained from a publicly available database (Borji & Itti, 2015). These groups of scenes differed with respect to bottom-up features (e.g. contrast, brightness), and top-down information (e.g. semantic knowledge regarding objects within the image) that are contained therein. The nature and time course of viewing are guided by both types of information, with bottom-up influences often preceding top-down influences (Henderson, Weeks, Phillip, & Hollingworth, 1999). For example, with respect to the influence of bottom-up information on viewing, research has shown that individuals make more fixations on regions of scenes that show physical discontinuity (Mackworth & Morandi, 1967) or less physically redundant stimuli (physical features that are less repetitive; e.g. irregular checkerboard, compared to a regular checkerboard; Berlyne, 1958). With respect to top-down cognitive influences, researchers have observed that individuals make more fixations, earlier during viewing, and for longer durations, on objects that do not fit with the semantic context of a given scene (Henderson et al., 1999; Loftus & Mackworth, 1978). Similarly, individuals also tend to make fixations on objects that are rated as easily recognizable on a future occasion (Mackworth & Morandi, 1967). Since bottom-up and top-down factors are known to elicit specific eye movements during scene viewing, we would expect the gaze signal patterns produced by eye movements to also vary with the bottom-up and top-down information present in and encoded from scenes.

In the current study, bottom-up information is indexed by Normalized Scan-path Saliency (NSS) scores provided in Borji et al. (2015; only NSS scores of scene categories computed by three
models were obtained from the web source associated with the paper; NSS scores for individual scenes are not provided). Similar to results documented in Borji & Itti (2015), the NSS scores obtained from the three models were averaged, to indicate the average power of these models in predicting the saliency of images based on bottom-up features (e.g. target size, color, intensity, orientation etc.; model details are provided in the literature review section of this thesis). The NSS score variation across scene categories suggests that bottom-up features and the information extracted from these features also might be variable across scene categories in the current dataset. This means MSE, which is a marker of the information encoded from scenes, also should vary across scene categories and track with NSS scores.

Next, top-down information relates to the semantic/cognitive information that is brought to bear upon viewing and may influence eye movements and resulting gaze signal patterns more than bottom-up information. Also, one would expect the semantic content present within scenes (e.g. depictions of people, places, objects) to vary across scene categories. Therefore, scenes that have more semantic content/relevance to viewing could be viewed differently from scenes that are, for instance, more abstract (arguably containing less semantic content). This then could also contribute to a change in gaze signal patterns across scene categories and a corresponding variation in the information encoded from scenes. Therefore, MSE is expected to vary across scene categories from a top-down point of view as well.

In comparing images that have more (high NSS) vs less (low NSS) bottom-up features and assuming that images with less bottom-up features have more top-down relevance, MSE is expected to differentiate the two types of images based on timescales. Specifically, based on the fact that viewing of bottom-up features precedes the viewing of top-down cues (Henderson, Weeks, Phillip, & Hollingworth, 1999), we could expect images with more bottom-up features to
have higher absolute entropy at finer timescales (corresponds to faster viewing) compared to images that have less bottom-up features. Conversely, images with less bottom-up features (and arguably, more top-down relevance) might have higher absolute entropy values at coarser timescales than images with more bottom-up features.

Finally, in order to test if the MSE metric would provide more information than what can be derived from static measures of eye movement such as the number of saccades (or fixations) and saccade amplitude, we 1) summarized the viewing behavior across the different scene types using these static measures, and 2) removed these measures from MSE to subsequently assess if MSE would still distinguish among scene types. If our gaze displacement signals are of the type 1/f, then MSE would be measuring the complexity of the entire signal trace that pre-dominantly consists of correlated/repeated displacement values over longer time intervals. Provided that MSE was initially able to distinguish among scene types based on their relative bottom-up relevance (NSS scores), we predicted that MSE would continue to index a contrast in scene type even after removing these static measures. This prediction was based on the fact that MSE measures the pattern/regularity of the entire signal trace, which is expected to change with the type of scene being viewed, whereas static measures only account for specific and discrete regions in this gaze signal pattern.
Chapter 3

3 Experimental study

3.1 Introduction

During viewing, saccadic eye movements move the fovea to different parts of the visual field (Henderson, 2003) to encode information from the regions that our eyes fixate on (Henderson, 2003). Previous research shows that we selectively view and encode visual information based on bottom-up and top-down factors that guide viewing (Antes, 1974; Henderson, 2003; Henderson, Weeks, Phillip, & Hollingworth, 1999; Loftus & Mackworth, 1978; Parker, 1978). Bottom-up factors relate to physical features of stimuli including color, texture, contrast, brightness and presence of edges/contours (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Reinagel & Zador, 1999; Henderson, 2003). Top-down influences relate to the cognitive/semantic information that is brought to bear on viewing behavior, including whether or not an object is semantically consistent within the context of a scene. In general, bottom-up factors influence eye movements (e.g. number of gaze fixations we make) before top-down cues over the time course of viewing (Henderson et al., 1999).

Eye movements reflect temporal and spatial variability in viewing – e.g. there’s the notion of where, as well as over what time duration we view objects. Methods that follow the principles of Non-linear Dynamic Systems (NDS) theory (used to describe, model and predict processes that change in a non-linear manner over time and/or due to various influences; Aks, 2008; Ogmen, Bruno & Breitmeyer, 2006; Gregson & Guastello, 2011) have been used by researchers to analyze how gaze signals produced by eye movements during viewing changes in relation to time (temporal dynamics) and space (spatial dynamics).
Previously, researchers have analyzed patterns of gaze signals recorded while individuals viewed various stimuli, including objects (Aks, Zelinsky, & Sprott, 2002; Fairbanks & Taylor, 2011) and scenes (Wu, Anderson, Bischof, & Kingstone, 2014). Research shows that gaze signal patterns contain eye fixations that are repeatedly directed towards the same image positions (Fairbanks & Taylor, 2011) over close as well as distant time intervals (Wu, Anderson, Bischof, & Kingstone, 2014; Aks, 2011), suggesting that gaze signals show viewing behavior that is similar over a range of time intervals as opposed to a particular time interval. Also, successive fixations may influence each other in non-proportional ways. These findings indicate that the temporal dynamics of eye movements can be better quantified by assessing gaze signals over multiple timescales as opposed to a single scale.

Similarly, previous analyses have highlighted changes in viewing patterns including changes in specific types of eye movements, with the spatial scales being analyzed. For instance, research done by Fairbanks & Taylor pertaining to visual search indicates that micro-saccades and saccades operate at different angular spatial scales (e.g. measured in degrees or pixels on the screen). Such changes in type of eye movements during viewing tend to recur over the course of viewing (Aks, 2008; Aks, Zelinsky, & Sprott, 2002; Fairbanks & Taylor, 2011). Evidence for the spatial dynamics of eye movements has also been put forward by Aks and colleagues (2002) in the context of visual search. In particular, they observed that when searching for a target object among distracters, gaze covered a greater vertical distance than horizontal distance over the time course of viewing (i.e. distance between y positions increased over time, whereas distance between x positions decreased over time). These results indicate that the degree of detail in the patterns of gaze signals changes across many spatial scales.
Collectively, the above evidence suggests that similar or correlated signals (e.g. repeated gaze fixations on the same image position) might be present in gaze signal patterns over multiple temporal and spatial scales (Aks, 2008; Aks et al., 2002; Wu et al., 2014; Aks, 2011; Fairbanks & Taylor, 2011), making gaze signal patterns complex in nature (Aks, 2008; Aks et al., 2002; Fairbanks & Taylor, 2011). This complexity places gaze signal patterns in a balanced state between perfectly ordered and completely chaotic/disordered signals (Aks, 2008; Aks et al., 2002; Fairbanks & Taylor, 2011). However, the exact type of complexity (signal dynamics or “color” of signals) tends to vary with the nature of gaze signals (e.g. absolute gaze position, direction of gaze position, displacement between successive gaze positions etc.), type of visual task during which eye movements were recorded (e.g. visual search task), and the resolution of the gaze signal data taken into consideration (e.g. sampling frequencies considered; Aks et al., 2002; Fairbanks & Taylor, 2011).

A particular type of gaze signal dynamics observed by Aks and colleagues (2002) called pink noise or $1/f$ type signals has gained interest among vision researchers (Fairbanks & Taylor, 2011) due to its nature and cognitive relevance to eye movements. Pink noise, by definition, contains highly correlated signals over longer time intervals and hence has higher signal amplitude or power (mean square of amplitude) at smaller frequencies (longer the time interval, lower the frequency; Fairbanks & Taylor, 2011). In Aks et al. (2002), researchers recorded eye movements while an individual searched for a visual icon (upright T) among distracters, after which they analyzed gaze signals produced from eye movements using various NDS methods. In particular, using a Fast Fourier Transform (FFT) analysis they deduced that gaze displacement signals (displacement between successive gaze positions) can be characterized as pink noise at lower frequencies, indicating the presence of similar gaze signals (e.g. repeated fixations) over longer time intervals. This then led to the idea that visual search behavior might not be random as
previously thought and that contingencies might exist across the fixations present in complex
gaze signal patterns (Aks et al., 2002). In particular, the presence of contingencies across
fixations over long time intervals (pink noise) indicates that a form of long-term spatial memory
might exist across these fixations (Aks et al., 2002). While the cognitive mechanism behind this
$1/f$ behavior is still unclear, they suggested that it could involve the selection of useful
information from a complex environment because $1/f$ signals tend to provide the optimal balance
between recovering information efficiently and making errors (Aks et al., 2002).

While the studies described above have explored the dynamics of gaze signals produced by eye
movements, most other studies have analyzed eye movement data from a “static” point of view.
These studies have used measures including number of gaze fixations (Chan, Kamino, Binns, &
Ryan, 2011; Heisz & Ryan, 2011; Henderson et al., 1999; Olsen et al., 2016; Parker, 1978;
Yeung, Ryan, Cowell, & Barense, 2013), number of saccades (Firestone, Turk-Browne, & Ryan,
2007; Loftus & Mackworth, 1978), amplitude of saccades (Antes, 1974; Henderson et al., 1999),
average viewing time analyses (Antes, 1974; Buswell, 1935; Firestone et al., 2007; Mackworth
& Morandi, 1967), mean fixation duration measures (Antes, 1974; Loftus & Mackworth, 1978)
and mean saccadic reaction time measures (Peltsch, Hemraj, Garcia, & Munoz, 2014). While
these measures provide a summary of how eye movements facilitate viewing, they do not
necessarily capture the manner in which complex gaze signal patterns produced by eye
movements unfold over time. In other words, these measures can be considered “static” because
they do not quantify the repeated signals (e.g. fixations) that might be present over longer time
intervals in gaze signal patterns (complexity of $1/f$ signals). Therefore, in an attempt to help
quantify the complexity of gaze signal patterns as they unfold over time, we have applied a
metric called Multi-Scale Entropy (MSE) on gaze signals.
MSE is a measure that quantifies the complexity of physiologic and physical signals of finite length by estimating the degree of signal irregularity, i.e. sample entropy, over multiple timescales (Costa, Goldberger, & Peng, 2005). The uniqueness of MSE, compared to other entropy-based measures, lies in its ability to assign higher entropy values to $1/f$ signals at longer timescales (indicative of longer time windows) compared to periodic signals (zero entropy) and white noise (contains no correlated structures at longer timescales, and is therefore not considered complex; Costa et al., 2005). Entropy has also been directly equated to the concept of information (Shannon, 1948; Guastello, 2011; Sarkar & Barat, 2008). Hence, entropy measures can quantify the information contained in signals (Guastello, 2011). Therefore, MSE, by assigning higher sample entropy values to structurally complex signals of the type $1/f$, essentially quantifies the meaningful information carried by correlated structures present over longer time intervals in these signals (Costa et al., 2005; Shannon, 1948).

We know from the work done by Aks and colleagues (2002) that gaze displacement signal patterns could contain low frequency signal regions that can be characterized as complex and informative $1/f$ signals. Based on this evidence along with previous applications of MSE on $1/f$ type signals (Costa et al., 2005), we predicted that if gaze displacement signals produced by eye movements pre-dominantly follow $1/f$ signal behavior, then MSE can be applied to quantify the complexity of these gaze signal patterns. Since MSE quantifies this complexity over multiple timescales, it can quantify how useful information is encoded from eye movements and contained in $1/f$ type gaze signal patterns as they unfold over time (i.e. over shorter and longer time intervals). Finally, since the bottom-up and top-down cues pertaining to different visual stimuli can guide eye movements in specific ways and hence, can differentially influence the information encoded from these visual stimuli, we expected the information (sample entropy) indexed by MSE to vary with the type of visual stimuli.
In order to test the above predictions, we applied MSE on gaze displacement signals produced by eye movements that were recorded while subjects freely viewed different types of scenes that belonged to 12 categories obtained from a database of scenes made available by Borji & Itti (2015). First, we analyzed if gaze displacement signals derived from the current task predominantly showed $1/f$ type behavior, using a Fast Fourier Transform (power spectrum) analysis. If these signals prove to be of the type $1/f$, then based on the application of MSE on simulated $1/f$ signals in Costa et al (2005), we predicted that MSE would assign high sample entropy values to gaze displacement signals at all timescales (indicative of both shorter and longer time windows).

Second, we predicted that MSE would differ on the basis of the information that can be encoded from scenes, i.e. different scene categories. In particular, in the scene dataset released by Borji & Itti (2015), scenes have been classified based on Normalized Scan-path Saliency (NSS) scores. NSS scores used in the current study indicate the predictive power averaged across three saliency models that predict the saliency of scenes based on bottom-up features (Borji & Itti, 2015; NSS scores of scene categories computed from three of the four models mentioned in the paper are available; NSS scores per scene are not provided). A lower NSS score assigned to a scene category would indicate that the bottom-up information contained in scenes belonging to that category is rather low, and the reverse is true for high NSS scores. Therefore, based on differences in information contained in different types of scenes, we expected MSE to track with the NSS of image categories. Also, assuming that images with less bottom-up features (low NSS) might have high top-down relevance, we expected that MSE would differentiate these images from images with high NSS scores based on timescales. Since bottom-up features are encoded faster/before top-down information can influence viewing (Henderson et al., 1999), we can assume that there will be a less of a need to repeatedly view a given image position over longer time intervals (view and the re-view after a long time period; corresponds to coarser timescales).
when encoding bottom-up features. Hence, we predicted that MSE of scenes that are predominantly bottom-up (higher NSS scores) will be higher at finer timescales compared to scenes that might have more top-down relevance. Meanwhile, scenes with high top-down relevance might have higher MSE than scenes that have more bottom-up features at coarser timescales.

In addition, we explored whether MSE distinguished among scenes even in instances where static measures such as the number of saccades (or fixations) do not. Moreover, the degree to which MSE is dependent on specific types (e.g. saccades) or extent (e.g. saccade amplitude) of eye movements is currently unknown and has been investigated in the current study. In order to address these questions, we controlled for basic eye movement metrics by independently regressing the number of saccades and average saccade amplitude values from sample entropy, to compute residuals. In this study, residuals are values that index the complexity/information remaining in gaze displacement signals, after removing the contribution of a specific basic eye movement metric (e.g. number of saccades). Provided that MSE distinguished among scene categories in the current task, if it is not largely dependent/based on any one basic static metric (number of saccades/fixations or average saccade amplitude), then the residuals computed after regressing these metrics from MSE should still distinguish among scene categories.

Overall, we expect the current work to verify that MSE can be used to quantify the complexity arising from repeated visual events (e.g. fixations) present over longer time intervals in gaze signals. This would serve to index the useful information encoded by eye movements and contained in these gaze displacement signal patterns as they unfold over time. We also hope that MSE would be able to differentially index the information encoded from different scenes based on their bottom-up/top-down information. If the results of our current study confirm our
predictions, this metric could be useful for detailing the complexity of information extraction that occurs over the course of viewing.

3.2 General materials and methods

3.2.1 Participants

Twenty healthy participants between the ages of 18 – 35 years (15 females; $M_{age}=22.95$, $SD_{age}=3.69$; $M_{education}=16.15$, $SD_{education}=2.41$) were recruited from the Rotman Research Institute volunteer pool in exchange for monetary compensation. All participants reported normal or corrected-to-normal vision at the time of testing (13/20 were tested on the visual acuity test and were proved to have normal vision; read contents of all 11 rows in the Snellen chart). They were not aware of the purpose of this experiment and they did not report ever having seen the experimental stimuli prior to the experiment. All experimental procedures were approved by the Research Ethics board of Baycrest Health Sciences.

3.2.2 Raw eye movement recordings

3.2.2.1 Stimuli

A total of 120 scenes were selected for the current experiment from a publicly available database of scene stimuli (CAT2000 dataset; Borji & Itti, 2015). This dataset contains scenes that were classified into 20 categories, of which 12 categories of 10 images per category were selected for our experiment. The categories that we selected were: 1) Art, 2) Fractal, 3) Inverted, 4) Jumbled, 5) Pattern, 6) Random, 7) Action, 8) Cartoon, 9) Indoor, 10) Outdoor Man-made, 11) Outdoor Natural, and 12) Social. These categories were selected in order to provide a broad range of meaning/semantic content to our experimental stimuli. The 10 images selected from each category for the current study were chosen because they came with saliency and fixation data
collected by Borji & Itti (2015) from their original experiment. Choosing the same images would then enable us, if necessary, to make comparisons between results from the current study and data collected by Borji & Itti (2015), in the future. As mentioned in the article that is associated with the public dataset (Borji & Itti, 2015) some of these categories elicit bottom-up cues more strongly (e.g. Pattern), while others elicit significant top-down cues (e.g. Social). Borji & Itti (2015) also recorded the average prediction power of three saliency models across scene categories, based on Normalized Scan-path Saliency (NSS) scores computed from each model per scene category. Scores for individual images are not provided. Since these models predict the saliency of images based on bottom-up features (Borji & Itti, 2015), the NSS scores could also serve as an approximate index of the extent of bottom-up contained in scenes that belong to a given category. We used the NSS per scene category computed from the three models to calculate an average NSS score per scene category (averaged across the three models) – this metric is used as an index of bottom up information for each scene category. Based on these average NSS scores, the 12 scene categories used in the current study can be rated in increasing order of NSS scores, i.e. extent of bottom-up information, in the following manner: Jumbled (0.63) < Outdoor Natural (0.68) < Pattern (0.71) < Cartoon (0.75) < Social (0.76) ~ Indoor (0.76) < Inverted (0.80) < Outdoor man-made (0.84) < Random (0.95) < Fractal (0.98) < Art (1.03) < Action (1.17). All images had resolution of 1920 x 1080 pixels, and this aspect ratio was maintained during stimulus presentation through the addition of black margins (see Figure 1). A brief description of the nature of each scene category, adapted from Borji & Itti, 2015, is presented next.

The Art category contains different styles of artworks and paintings, both computer-generated and man-made. The Fractal category contains artificial images from different kinds of fractals. The Inverted category consists of scenes that are upside down. The Jumbled category was built
by Borji & Itti (2015) by randomly dividing each scene into partitions of specific dimensions. The length and width of these partitions were randomly chosen from numbers: 2, 3, 4 and 5. These partitions were then randomly shuffled in the creation of jumbled scenes. The Pattern category consists of psychological patterns that are commonly used in bottom-up saliency model evaluation in studies, including pop-out and conjunction. The authors included texture defects as well. The Random category contains images captured from random viewpoints and orientations, which accounts for the issue of individuals looking at the center of the screen as a viewing strategy in laboratory eye tracking setups. The Action category contains scenes that the authors obtained from the Stanford action dataset (Yao et al., 2011). The Cartoon category contains computer generated scenes that are non-natural and simple in nature. The Indoor and Outdoor Man-made categories consist of scenes that the authors obtained from the SUN database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). Of the two categories of outdoor scenes, the Outdoor Man-made category consists of structures built in open space, including bridges and roads. The Outdoor Natural category is made up of natural scenes (e.g. mountains, forests etc.). Finally, the Social category contains images of people having social interactions and the authors focused on preserving semantic attention cues such as gaze direction in these scenes. Figure 1 depicts sample images taken from each scene category.
Figure 1. Sample images taken from the 12 scene categories selected for the current experiment are depicted in this figure. Scenes were selected from the dataset that has been made publicly available by Borji & Itti (2015).

3.2.2.2 Experimental design

The 120 scenes that were selected from the publicly available database (Borji & Itti, 2015) belonged to one of 12 scene categories, resulting in 10 scenes per category. These scenes were randomly presented to participants without respect to category. The duration of scene presentation employed in the original experiment conducted by Borji & Itti (2015) was replicated in our experiment. Specifically, each scene was presented for 5 seconds, followed by a 3 second gray screen. Participants were instructed to view the scenes freely without any restrictions.

3.2.2.3 Apparatus and calibration criteria

Scenes were presented to participants on a 24-inch LCD monitor (resolution: 1920 x 1080) from a distance of 70cm, such that the monitor subtended approximately 24.03 degrees of visual angle based on the height of the screen (vertical/Y plane). A chin rest was used to stabilize head movements and maintain consistent viewing distance within and across subjects. Monocular eye movements of participants were recorded using an EyeLink II eye tracker at a sampling rate of
500Hz with a spatial resolution of 0.01° and a typical average accuracy of <0.5° (SR Research Ltd., Mississauga, ON, Canada). An initial 9-point calibration was done at the beginning of the experiment, and drift corrections were done before the start of each trial. The initial calibration was repeated until the average gaze error was below 0.5° and the maximum gaze error was below 1°.

3.2.3 Analysis of Visual Sampling Behavior

Using the Data Viewer program associated with the EyeLink II eye tracking system (SR Research Ltd., Mississauga, ON, Canada), we characterized overall viewing to the scenes with the following measures: the number of fixations, the number of saccades, and the average saccade amplitude (angular displacement of transitional eye movements, in degrees). A one-way ANOVA with an alpha of 0.05 was used to assess potential statistical differences in these basic metrics across scene categories.

Note that the duration of trial over which these measures were obtained was 4960msec. from the start of the image. This was done in order to match the trial duration used for subsequent MSE calculations. In order to maintain consistency between the number of events (e.g. fixations, saccades etc.) calculated by the Data Viewer program versus the number of events that fell within the interested period of 4960msec from scene onset, calculated based on text records of visual events obtained from the Visual EDF2ASC converter (SR Research Ltd., Mississauga, ON, Canada), we hid any event that spanned the start of this period (i.e. the scene onset boundary) on Data Viewer before obtaining measures of eye movements. This was done because Data Viewer, by default, includes the event (usually an event marking the start of a fixation) that occurs just before the scene onset boundary in the total number of events (e.g. number of fixations) calculated as being present within a defined period of interest. By hiding the event that
spans the start of the period of interest and limiting the interested time period to 4960msec from scene onset, we ensured that only events that mark the start of fixations and start of saccades contribute to the “number” of fixations and saccades, respectively, both by Data Viewer and the number that can be computed based on raw text recording of visual events. Therefore, the number of fixations in the current dataset within defined period of interest (scene onset to 4960 msec.) is not always 1 less than the number of saccades (as per the conventional relationship between the number of fixations and saccades), but in some cases is equal to the number of saccades.

3.2.4 Eye movement data reduction (calculation of angular displacement)

Participant specific raw eye movement position data (x and y positions in pixels) for each trial was reduced to angular displacement signals (visual angle in degrees) over a duration of 4960msec in order to depict the occurrence of saccades. A sliding window approach was taken to calculate angular distance with each window consisting of 15 data points (corresponding to 30msec of data). Since our sampling rate is rather high, this procedure enabled us to filter in some of the high frequency noisy signals to avoid the decimation of saccades. We used a 30 msec. time window to be lenient with the noisy signals. Signals that corresponded to the up-phase or down-phase of blinks, or empty regions that were found in the sequence of position data, which could include blinks themselves, were eliminated by zeroing 15 data points (30 msec.) on either side of empty regions.

3.2.5 Fast Fourier Transform analysis

In order to identify the signal composition or “color” of gaze displacement signals, we computed the discrete Fourier transform (DFT) of gaze displacement time series using the built-in Fast Fourier Transform (FFT) algorithm in the MATLAB signal processing toolbox (Version 6.16;
R2011b): \( Y = \text{fft}(X,n) \). In this algorithm, ‘X’ refers to the trial-specific gaze displacement signal matrix and ‘n’ refers to the input length, which in this case was taken to be the next power of 2 from the original signal length (which pads the signal with trailing zeros) to increase the performance of the ‘fft’ function (https://www.mathworks.com/help/matlab/ref/fft.html).

The DFT of gaze displacement time series was then converted to the frequency domain to create a power spectrum (signal power versus frequency) in order to identify the signal composition of these signals. The MATLAB function: \( P = \text{abs}(Y/n) \) was used to compute signal power (P). Similarly, the frequencies (f) were derived from the function: \( f = \text{Fs}*(0:(n/2))/n \), where ‘Fs’ refers to the sampling frequency (i.e. 500Hz in the current study).

Finally, the power spectrum was converted to a log-log plot (log of power vs log of frequency) in order to perform a regression analysis on this data form. The gradient (size of which is equal to the power law exponent: \( \alpha \)) of the regression line applied on the linear region of signals (this region is considered to show recurrent patterns over multiple temporal scales, and hence shows power law scaling behavior; Fairbanks & Taylor, 2011) would indicate the type of signal dynamics (“color”) followed by these gaze displacement signals (Aks et al., 2002). In particular, we were interested in identifying whether these signals can pre-dominantly be characterized by pink \((1/f)\) power laws because based on previous research MSE is known to be able to quantify the irregularity and information contained in \(1/f\) type signals relative to periodic signals or meaningless white noise (Costa et al., 2005). This can be verified by assessing if the magnitude of the regression gradient falls between 0 and 2, as this gradient \((\alpha)\) range is commonly known to be indicative of pink power laws in most physical signals (Fairbanks & Taylor, 2011). This analysis would indicate the presence of correlated signal structures occurring over longer time intervals, making these signals interesting and meaningful. It would also then provide
justification to use MSE to quantify the structural complexity and useful information contained
in these gaze displacement signals.

3.2.6 Calculation of Multi-Scale Entropy (MSE) from angular displacement
signals

In order to capture the changes in visual events (e.g. saccades) that occur at different timescales,
multiple coarse-grained eye movement “time series” were created from the original angular
displacement time series by progressively down-sampling the angular displacement time series
\{x_1, \ldots, x_i, \ldots, x_n\} by trial. The original angular displacement time series consisted of data points
that occurred at 2msec time intervals over the duration of a given trial. For a given timescale, \( \tau \),
the coarse-grained time series \( \{Y^{(\tau)}\} \) was constructed by averaging the angular displacement
signals that fell within non-overlapping windows of timescale length (i.e. \( \tau \)). Each element of a
given coarse-grained time series, \( j \), was calculated according to equation 1 below (Costa et al.,
2005):

\[
Y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \frac{N}{\tau} = N^{(\tau)}.
\]

**Equation 1** – Calculation of each element of a given coarse-grained timeseries

Down-sampling results in a change in frequency of timeseries (the timeseries appears to have
higher frequency at finer vs coarser timescales). This down-sampling procedure was used to
create 80 timescales. Down-sampling is important to create multiple timeseries at various scales
than just the shortest one. MSE can then assign entropy values to all these timeseries and
highlight the timeseries (timescales) that are truly complex (have the highest entropy). In this
manner down-sampling makes MSE a more powerful tool than just sample entropy, which
captures the shortest scale only and can assign higher entropy to chaotic vs complex signals. Figure 2 (top panel) illustrates the nature of gaze angular displacement (degrees) as a function of time (msec.), for a sample trial, at finer (e.g. timescale = 1, 30; Figure 2) and coarser (e.g. timescale = 60; Figure 2) timescales. Subsequently, the sequence of angular displacement signals corresponding to each coarse-grained time series was analyzed using the information theoretic tool, MSE.

Once the coarse-grained time series were created, sample entropy ($S_E$) was calculated from angular displacement signals, at each timescale. Figure 2 (bottom panel) depicts $S_E$ values calculated from corresponding gaze angular displacement signals at finer (e.g. timescale = 1, 30) and coarser (timescale = 60) timescales (series of angular displacement signals that correspond to each timescale are shown in the top panel). Sample entropy at each timescale was calculated using equation 2 below:

\[
S_E(m, r, N) = \ln \frac{\sum_{i=1}^{N-m} n_i^m}{\sum_{i=1}^{N-m} n_i^{m+1}},
\]

**Equation 2 – Calculation of sample entropy at each timescale**

where $m$ is the pattern length, $r$ is the similarity criterion, $N$ is the length of a given time series and $n_i^m$ is the number of matches. Sample entropy quantifies the predictability of signals along the length of the time series ($N$). The pattern length ($m$) was set to 2, which means that the predictability of angular displacement at each time series was assessed by considering the patterns of angular displacement signals in sequences of 2 versus 3 consecutive data points.
respectively. The similarity criterion \((r)\) was set to 0.5, meaning that two angular displacement signals were considered to match if the absolute difference between the values was less than or equal to 50\% of the time series standard deviation. These parameters were set based on the guidelines published in Richman and Moorman (2000). These guidelines are based on the application of sample entropy on other biological signals (e.g. clinical cardiovascular time series). For each trial, sample entropy was calculated in this manner at each timescale (1-80), thus giving rise to MSE. A trial average of MSE was also calculated for each scene category and participant.

Figure 2 - The top panel illustrates the nature of gaze angular displacement as a function of time, calculated for a sample trial (an art scene viewed by a participant) and captured at finer (timescale = 1, 30) and coarser (timescale = 60) timescales. The bottom panel depicts sample entropy calculated from angular displacement signal patterns that correspond to timescales 1, 30 and 60.
3.2.7 Controlling for visual sampling behavior

3.2.7.1 Calculating residuals after regressing number of saccades from MSE

In order to determine whether MSE is merely reflective only of the number of saccades, as opposed to the entire and continuous pattern exhibited by gaze displacement signals, the number of saccades made on stimuli were regressed from the sample entropy values at each timescale for each trial in order to obtain residuals that were subsequently statistically assessed for reliable differences across scene categories. If MSE quantifies the complexity/meaningfulness of continuous gaze displacement patterns beyond the number of saccades alone, then MSE should reveal differences across scene conditions even after removing the number of saccades participants made on these scenes. Note that the number of gaze fixations in the current dataset is either be equal to or 1 less than the number of saccades. Hence, a similar output is expected from regressing the number of gaze fixations from MSE.

3.2.7.2 Calculating residuals after regressing average saccade amplitude from MSE

Based on our preliminary exploration of the eye movement data, we observed that the average extent (amplitude) of saccades, varied across scene categories. For instance, we observed that on average, individuals made longer saccades when viewing jumbled scenes (5.58 degrees), compared to random scenes (4.44 degrees; t=3.44; p<0.002). Hence, in order to identify whether potential differences across scene categories as indexed by MSE could be attributed to differences in saccade amplitude, we regressed the average saccade amplitude from sample entropy for each trial at each timescale. If MSE is not dependent on the saccade amplitude alone,
we would expect it to index a significant and reliable contrast across scene categories even in the absence of saccade amplitude values (after regressing saccade amplitude from MSE).

3.2.8 Statistical assessment through Partial Least Squares (PLS)

The statistical significance of distributed MSE data as well as residuals obtained from the regression analyses across participants and scene categories was assessed using Partial Least Squares (PLS). PLS is a multivariate technique that is used to identify changes in distributed signals in relation to the experimental manipulation/task demands (task PLS) or task performance (behavioral PLS; McIntosh & Lobaugh, 2004).

3.2.8.1 Task PLS

The data matrix created in preparation for PLS analysis was organized in the following general format: [participants x scene category] in rows and [timescale] in columns, with each cell consisting of a MSE or residual value. The original data matrix was first mean-centered to highlight scene category differences (the grand mean calculated across scene categories was subtracted from individual sample entropy values for each timescale). The mean-centered data matrix then underwent singular value decomposition, which projected the data matrix onto a set of orthogonal latent variables (LV). Each LV consists of three components: 1) singular value, 2) v-vector or design salience, which represents the experimental design profile/contrasts that best characterize the original data matrix, and 3) u-vector, which represents the temporal distribution (timescales) of the original MSE data matrix (McIntosh & Lobaugh, 2004).

3.2.8.2 Statistical assessment of LVs

The LVs derived from the task PLS analysis was statistically assessed in two ways. First, the degree to which the effects of specific LVs were statistically different from chance was assessed
through 500 permutation tests. Permutation tests involve randomly assigning conditions for each subject via resampling without replacement, and subsequently recalculating PLS for each new sample (McIntosh & Lobaugh, 2004). The significance values of LVs were calculated from the proportion of times permuted singular values exceeded observed singular values (McIntosh & Lobaugh, 2004).

Second, the stability of the LVs was assessed through the bootstrap estimation of standard errors, i.e. bootstrap ratio. The bootstrap procedure involves resampling with replacement in which condition assignments are fixed, after which PLS is recalculated for each bootstrap sample (McIntosh & Lobaugh, 2004). In the current study, 250 bootstrap samples were produced to assess the reliability of the temporal distribution (timescales) of MSE or residuals, across subjects. The bootstrap ratio is calculated as the ratio of the salience of a given bootstrap sample to the standard error of the generated bootstrap distribution (McIntosh & Lobaugh, 2004) and is roughly equivalent to a z-score when the bootstrap distribution is normal (Efron & Tibshirani, 1986). A bootstrap ratio of 2, which is roughly equivalent to a 95% confidence interval, was chosen as the threshold in this study. Accordingly, for a given LV, the timescales that had bootstrap ratios greater than 2 were indicative of the temporal distribution that was reliable across subjects.

3.3 Results

3.3.1 Visual Sampling Behavior

Our initial data exploration revealed that across all scene categories, subjects made 14.23 fixations on average. Similarly, the average saccade count across all scene categories and subjects was 14.34, with the mean number of saccades made by subjects when viewing certain
categories (maximum: cartoon scenes; 15.52 saccades) appearing to be slightly greater than others (minimum: fractal scenes; 13.43 saccades; t=2.48; p=0.02). Moreover, on average, the mean saccade amplitude (extent of saccades) was 4.96 degrees, across all subjects and scenes with participants, on average, making longer saccades while viewing jumbled scenes (maximum; 5.58 degrees) compared to random scenes (minimum; 4.44 degrees; t=3.44; p=0.002).

In order to identify whether visual sampling differed across the 12 scene categories, we calculated the average number of fixations, average number of saccades and average saccade amplitude for each type of scene (Table 1). A one-way ANOVA showed that the number of fixations ($F_{(11,228)}=1.13$, $p=0.34$) and saccades ($F_{(11,228)}=1.11$, $p=0.35$) did not significantly differ across scene conditions. However, a significant difference in saccade amplitude was seen across scene conditions ($F_{(11,228)}=2.37$, $p=0.009$).

<table>
<thead>
<tr>
<th>Scene category</th>
<th>Eye movement measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Fixations</td>
</tr>
<tr>
<td>Art</td>
<td>14.60 (2.39)</td>
</tr>
<tr>
<td>Fractal</td>
<td>13.30 (2.88)</td>
</tr>
<tr>
<td>Inverted</td>
<td>13.72 (3.00)</td>
</tr>
<tr>
<td>Jumbled</td>
<td>14.76 (2.51)</td>
</tr>
<tr>
<td>Pattern</td>
<td>14.26 (2.90)</td>
</tr>
<tr>
<td>Random</td>
<td>13.73 (2.37)</td>
</tr>
<tr>
<td>Action</td>
<td>13.84 (2.81)</td>
</tr>
<tr>
<td>Cartoon</td>
<td>15.43 (2.47)</td>
</tr>
<tr>
<td>Indoor</td>
<td>14.72 (2.97)</td>
</tr>
<tr>
<td>Scene Category</td>
<td>Mean (SD) 1</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Outdoor Man-made</td>
<td>13.65 (3.01)</td>
</tr>
<tr>
<td>Outdoor Natural</td>
<td>13.76 (3.25)</td>
</tr>
<tr>
<td>Social</td>
<td>15.03 (2.69)</td>
</tr>
</tbody>
</table>

Table 1 – The table depicts visual sampling behavior for each scene category. Standard deviations are indicated in parentheses. The asterisk indicates a significant main effect of scene type on the average saccade amplitude.

3.3.2 Signal composition of gaze displacement signal time series

Results from the discrete Fourier transform (DFT) and subsequent power spectrum analysis (figure 3a and log-log plot in figure 3b) showed that these signals had higher power at lower signal frequencies (inversely, longer time intervals). Furthermore, a regression analysis done on the linear signal region of the log-log plot (log (frequency) > 1; figure 3c) produced a coefficient (gradient) of -1.36, indicating $1/f$ signal power. Note that figures 3a-c are examples depicting the nature of gaze displacement signals derived from one subject and trial (in this example, an art scene).
Figure 3 – (a) The figure depicts a power spectrum of signal power versus frequency, obtained from a discrete Fourier Transform (DFT) applied on gaze displacement signals derived from a subject and trial (art scene). The signals show higher power at lower frequencies. (b) The plot depicts a power spectrum of log transformation of the power spectrum (log-log plot of power versus frequency). (c) The plot depicts the linear high frequency signal region (log (frequency) > 1) of the log-log plot of the power spectrum. The linear regression line that fits the signal trace best is depicted in red. The gradient of the regression line was -1.36.

3.3.3 Eye movement based MSE

Subject specific MSE was calculated from angular displacement signals for each trial and then averaged across trials and subjects to obtain a grand mean MSE (+/- SE) for each scene category (Figure 4: a-l). As depicted in Figure 3, across all 12 scene categories, MSE profiles show a
similar change in sample entropy over different timescales (τ). Specifically, sample entropy progressively increases over $1 \leq \tau \leq 30$, followed by a decrease in the slope of MSE curves over $30 < \tau \leq 40$, after which the curves plateau over $\tau > 40$. Essentially, sample entropy peaks around $\tau = 40$. A visual representation of the manner in which the MSE profiles of each type of scene relate to each other is provided in figure 5.
Figure 4 – The figure depicts the average MSE (+/- SE) profiles for each scene category: (a) Art, (b) Fractal, (c) Inverted, (d) Jumbled, (e) Pattern, (f) Random, (g) Action, (h) Cartoon, (i) Indoor, (j) Outdoor Man-made, (k) Outdoor Natural and (l) Social. MSE was calculated by averaging sample entropy as a function of timescales, across trials and subjects, for each scene category. For all scene categories, MSE increases over finer timescales (1 ≤ τ ≤ 30) and plateaus at relatively coarser timescales (τ > 40). The peak in sample entropy is around τ = 40.

Figure 5 – Average MSE profiles of all 12 scene categories are presented in the same plot in this figure. Standard error bars have been removed from the MSE profiles for clarity. On average, jumbled scenes have the highest MSE, while action scenes have the lowest, and this difference is pronounced over τ > 40.

The contrast in MSE between the scene categories across all subjects was statistically assessed using a data driven task PLS. Permutation tests revealed two significant LVs. The 1st LV showed a distinct contrast in MSE across the 12 scene categories (p < 0.002 figure 6a). Specifically, based on the degree of variability shown by error bars in the task PLS plot depicted in figure 6a, it is evident that this contrast was driven by significant differences in MSE that existed across fractal
(category label 2; figure 6a), jumbled (category label 4; figure 6a), random (category label 6; figure 6a), action (category label 7; figure 6a), cartoon (category label 8; figure 6a), indoor (category label 9; figure 6a) and outdoor natural scenes (category label 11; figure 6a). Taken together with differences in MSE profiles that were seen across these 7 scene categories (depicted in figure 5), it is evident that MSE was significantly higher when individuals viewed jumbled, cartoon and indoor scenes compared to the remaining scenes (i.e. fractal, random, action and outdoor natural scenes). Moreover, as depicted in the bootstrap plot shown in Figure 6b, the contrast highlighted by the task PLS bar chart in figure 6a appears to be reliable for \( \tau \geq 5 \), indicating that this contrast is evident across most of the timescales. Note that increasing the bootstrap ratio threshold to 3, does not drastically change the reliable range of timescales. The contrast shown by LV 2, while significant, accounted for less than 9\% of the variance and therefore has not been reported in detail in this paper.

**Figure 6** – **(a)** The plot depicts the data-driven contrast between scene category specific MSE values that was significantly expressed across the timescales, as determined by permutation tests. This contrast was derived from the 1\textsuperscript{st} LV. Error bars denote the 95\% confidence intervals derived from the upper and lower percentiles of the bootstrap distribution. The u-score average refers to the mean-centered effect of the contrast (scores that represent the u-vector distribution across subjects and scene categories have been mean-centered on subject). **(b)** The plot depicts
the bootstrap distribution, which highlights the temporal (timescale-based) reliability of the contrast depicted in figure 6(a), across all subjects. The threshold for reliability was set to an absolute bootstrap ratio of 2.

3.3.4 Analysis of MSE controlling for visual sampling behavior

3.3.4.1 Regressing the number of saccades from MSE

In order to comprehensively account for the influence of the number of saccades on MSE across all trials and subjects, the number of saccades was regressed from sample entropy values by trial, for a given timescale and subject. The resulting residuals were averaged across trials, and subsequently assessed for potential contrasts across subjects and scene categories using a task PLS. Permutation tests revealed 1 significant LV (p <0.002), which indicated that the contrast across scene categories was above chance (Figure 7a). As depicted in figure 7a, based on the variability indexed by error bars, this contrast is driven by art (category label 1; figure 7a), random (category label 6; figure 7a) and cartoon scenes (category label 8; figure 7a). This contrast was reliable across subjects over all timescales (Figure 7b; BSRs exceeded the threshold of 2). Since BRSs are negative, the art, random and cartoon scenes are considered to have lower entropy. Note that increasing the bootstrap ratio threshold to 3 eliminates the reliability of very fine (e.g. τ < 15) and very course timescales (e.g. τ > 75).

1 The same pattern of results was observed when the number of gaze fixations was regressed from MSE. This contrast was also significant (p=0.002).
Figure 7 – (a) The plot depicts the data-driven contrast in residuals between scene categories that was significantly expressed across the timescales, as determined by permutation tests. The number of saccades was regressed from MSE to produce the residuals. Error bars denote the 95% confidence intervals derived from the upper and lower percentiles of the bootstrap distribution. The u-score average refers to the mean-centered effect of the contrast (scores that represent the u-vector distribution across subjects and scene categories have been mean-centered on subject).

(b) The plot depicts the bootstrap distribution, which highlights the temporal (timescale-based) reliability of the contrast depicted in figure 7(a), across all subjects. The threshold for reliability was set to an absolute bootstrap ratio of 2.

3.3.4.2 Regressing the average saccade amplitude from MSE

Differences in the average saccade amplitude were observed across scene categories.

Accordingly, in a final regression analysis, we removed the average saccade amplitude for a given trial from the corresponding sample entropy value, at each timescale, for each subject, to identify if the complexity of eye movement signals could differ across scene conditions even after removing saccade amplitude from the signals. Similar to the analyses done in section 4.1 above, we averaged the residuals that resulted from this analysis, across trials, and analyzed the contrast across subjects and scene categories using a task PLS analysis. Accordingly,
permutation tests revealed a significant LV (p=0.006; figure 8a) indicating a significant contrast in residuals across scene categories. Based on the variability indexed by error bars in figure 8a, this contrast appears to be mainly driven by outdoor natural scenes and the entropy is of this category is low (category label 11; figure 8a). This contrast was reliable across subjects over all timescales (figure 8b). Note that increasing the bootstrap ratio threshold to 3 limits the reliability to mostly finer timescales (e.g. \( \tau < 30 \)).

**Figure 8** – (a) The plot depicts the data-driven contrast in residuals between scene categories that was significantly expressed across the timescales, as determined by permutation tests. The average saccade amplitude was regressed from MSE to produce these residuals. Error bars denote the 95% confidence intervals derived from the upper and lower percentiles of the bootstrap distribution. The u-score average refers to the mean-centered effect of the contrast (scores that represent the u-vector distribution across subjects and scene categories have been mean-centered on subject). (b) The plot depicts the bootstrap distribution, which highlights the temporal (timescale-based) reliability of the contrast depicted in figure 8(a), across all subjects. The threshold for reliability was set to an absolute bootstrap ratio of 2.
3.3.5 Post-hoc analyses

3.3.5.1 Correlation between peak MSE and saccade amplitude across scene categories

Although the results of regressing saccade amplitude from MSE showed a significant contrast across scene categories, this effect was mainly driven by only one scene category (outdoor natural scenes). This result suggests that MSE might be associated with saccade amplitude. Hence, we analyzed the linear relationship between the peak (maximum) MSE (MSE by trial, averaged across subjects) assessed over all timescales and average saccade amplitude (by trial, averaged across subjects), across all scene categories, to statistically gauge how much of MSE can be accounted for by the extent of saccades present in eye movement signals. A simple correlation of both metrics is depicted in figure 9. Results from a simple regression analysis, with an alpha of 0.05 revealed a significant positive correlation between both metrics ($R^2=0.12$, $F(1,118)=15.76$, $p<0.001$).

![Figure 9](image-url)

**Figure 9** – The plot depicts a significant positive correlation between the peak MSE, and saccade amplitude, both represented by trial, across the12 scene categories.
3.3.5.2 Comparison between average NSS scores by scene category and MSE

Upon comparing the average NSS scores (averaged across three saliency models) and peak MSE (averaged across trials and subjects) assessed over all timescales, by scene category, it is evident that these measures have a significant negative correlation between them ($R^2=0.42$, $F(1,10)=7.22$, $p=0.02$). This is depicted in figure 10. In other words, scene categories with low NSS (e.g. Jumbled) were assigned higher entropy by MSE. Similarly, scene categories with high NSS (e.g. Action) were assigned low entropy by MSE.

![Figure 10](image)

**Figure 10** – The plot shows a significant negative correlation between peak MSE and average NSS score across the 12 scene categories

3.3.5.3 Comparison between average NSS scores and basic EM metrics

When analyzing the relationship between NSS scores and the basic EM metrics: number of fixations, saccades and average saccade amplitude, only the saccade amplitude had a significant negative correlation with average NSS scores (Figure 11; $R^2=0.39$, $F(1,10)=6.32$, $p=0.03$)
across the 12 scene categories. Number of fixations ($R^2=0.16$, $F(1,10)=1.95$, $p=0.19$) and saccades ($R^2=0.17$, $F(1,10)=2.01$, $p=0.19$) did not significantly relate to NSS scores.

![Figure 11](image)

**Figure 11** – The plot depicts a significant negative correlation between the average saccade amplitude measure and average NSS score, both represented by scene category.

### 3.4 Discussion

In our current work, we applied a signal complexity measure known as Multi-Scale Entropy (MSE) on gaze displacement signals in order to assess how the complexity of these signals, resulting from the repetition of signals (e.g. fixations) over multiple temporal and spatial scales (Aks, 2008; Aks et al., 2002; Wu et al., 2014; Aks, 2011; Fairbanks & Taylor, 2011) and information contained therein (Costa et al., 2005; Shannon, 1948) changes as viewing unfolds over time. Eye movement signals were recorded while participants freely viewed various natural and artificial scenes obtained from Borji & Itti (2015).

In this study, we tested two main hypotheses. First, provided that our gaze displacement signals can be characterized as $1/f$ signals containing repeated signal events at longer time intervals, we
hypothesized that MSE would assign high sample entropy values to these signals at all timescales. This hypothesis was based on the fact that repeated signals over longer timescales in 1/f gaze signals implies that contingencies exist across fixations, which might be indicative of a form of long term spatial memory involved with the encoding of useful information from the environment (Aks et al., 2002). We predicted that MSE would highlight this information through assigning high sample entropy to gaze signals. Based on Costa et al (2005), MSE assigned high entropy to simulated 1/f signals at all timescales but distinguished 1/f from white noise mostly at coarser timescales. Based on this result, we predicted that MSE would assign high sample entropy values at all timescales in our study as well. Second, we hypothesized that MSE would vary across scene categories. The bottom-up information contained in scene categories selected for the current study varied based on the Normalized Scan path Saliency (NSS) scores assigned to scene categories in Borji & Itti (2015; higher the NSS score, higher the bottom-up content). Similarly, the top-down content of scenes is expected to vary across scene categories as well (e.g. scenes depicting social interactions might have higher top-down value than images of patterns). Since both bottom-up and top-down cues guide viewing in specific ways (Henderson et al., 1999; Loftus & Mackworth, 1978), we predicted that the complexity of gaze signal patterns produced by eye movements and information encoded from scenes will also differ across scene categories and will be differentially indexed by MSE. In relating bottom-up with top-down content, one interpretation is that scene categories with lesser bottom-up content (indexed by low NSS scores) might have higher top-down relevance. Since we tend to view bottom-up features quicker than the time taken for top-down cues to affect viewing (Henderson et al., 1999) we predicted that MSE would be higher at shorter timescales for scenes that had higher bottom-up value (high NSS score) relative to scenes with low NSS score, as shorter timescales correspond
to faster viewing. Similarly, we predicted that scenes with low NSS scores (possibly, higher top-down relevance) will have higher MSE than scenes with high NSS scores at coarser timescales.

Prior to testing our hypotheses, we identified the pre-dominant signal behavior exhibited by our gaze displacement signals using a Fast Fourier Transform (FFT) analysis. A log-log plot of the power spectrum produced a regression line gradient of 1.36. Since the corresponding gradient of $1/f$ signals usually falls between 0 and 2 for most physical signals (Fairbanks & Taylor, 2011), this result suggests that our gaze displacement signal is of the type $1/f$ signals at most sampling frequencies, indicating that in theory, these signals contain highly correlated or similar signal structures (e.g. repeated gaze displacement values) at longer time intervals. This indicates that similar to the gaze displacement signal behavior presented in Aks et al. (2002), there might be contingencies existing across fixations over longer time intervals. The information encoded during these fixations could be contributing to a form of spatial memory that exists across these fixations (Aks et al., 2002). Based on previous applications of MSE on $1/f$ type signals, this result justifies applying MSE on these signals to quantify this useful information encoded by eye movements and contained in these complex signals.

In applying MSE on gaze displacement signals to test our first hypothesis, we found that sample entropy progressively increased with timescale and plateaued over $\tau > 40$. This trend was seen across all scene types. When observing the gaze displacement versus time profiles of trials across all timescales (e.g. fig 2 in general methods), it is evident that in the current study, the number of saccadic visual events only really changes (decreases) at around $\tau = 60$ (note that the changes in MSE at this timescale differs across scene categories). However, the signal amplitude appears to progressively decrease with timescale. This decrease in amplitude with timescale is an expected consequence of the coarse-graining procedure. In other words, high amplitude signals (captured
at shorter timescales) are almost periodic and this periodicity is filtered out by the coarse-graining procedure, resulting in more irregular time series characterized by smaller signal amplitudes at longer timescales (Costa et al., 2005). Hence, in the current study, timescales are classified as coarser vs finer scales based largely on when an obvious change (decrease) in the number of saccadic visual events is seen during the coarse-graining procedure. Accordingly, it is safe to assume that $\tau \geq 60$ are coarser scales and $\tau < 60$ are finer scales. Since the current results show that MSE remains high over $\tau > 40$, we can infer that MSE assigned consistently high sample entropy values to both finer ($40 < \tau < 60$) and coarser ($\tau \geq 60$) timescales. This suggests that the maximum amount of information is encoded with the help of saccadic visual events occurring both closer together (e.g. with the time interval between them being in the range: 80msec. -120 msec. based on an eye tracker sampling rate of 500Hz) and further apart in time (e.g. with the time interval between them being greater than 120msec. based on an eye tracker sampling rate of 500Hz). This then is consistent with results from Costa et al., 2005, where they applied MSE on simulated $1/f$ signals and observed a high and stable MSE (maximum information) at both finer and coarser timescales.

However, in the current study, MSE also indexes an increase in sample entropy over $1 < \tau \leq 40$ before the plateau region of $\tau > 40$, indicating that saccadic visual events that occur very close together in time (finer timescale regions; with the time interval between them being less than 80 msec. based on an eye tracker sampling rate of 500Hz) also contribute to information encoding to some extent, even if they do not help encode the maximum amount of information. This result is consistent with previous evidence of a high correlation seen between consecutive eye fixations (Aks et al., 2002) implying that a given visual event (e.g. fixation) predicts the occurrence of the subsequent visual event. This predictability seen between consecutive visual events could explain the non-zero sample entropy values assigned by MSE to finer timescales before the
plateau region ($\tau \leq 40$). Very fine timescales (e.g. $\tau < 5$) have low MSE that does not index reliable scene differences across subjects (BSR < 2 in figure 6B), possibly because highly frequent and irregular eye movement jitters are captured in this range, and they are not informative for scene differentiation from a cognitive perspective but may reflect physiological operations (e.g. stabilization of image on retina).

In addressing our 2nd hypothesis, we observed that MSE significantly differed across scene types. In particular, scenes that consisted of a jumbled assortment of sections of a given scene had the highest MSE compared to, for instance, scenes that were captured from random viewpoints. Interestingly, based on the average NSS scores calculated for scene categories from the data obtained from the Borji & Itti (2015) dataset, jumbled scenes had the lowest NSS score, while the action scene category had the highest, indicating that saliency models were able to use bottom-up cues to predict jumbled scenes to a lesser extent compared to action scenes. One possible interpretation of NSS scores is that scenes that contain fewer bottom-up cues (low NSS scores) might have higher top-down relevance. Based on this interpretation, we can assume that jumbled scenes contained more semantic information and/or required more top-down cognitive operations than random scenes in the current study. This could be because parts of a given scene that are put together in a jumbled manner might be semantically ambiguous upon initial viewing, and hence top down information that is brought to bear upon viewing is required to interpret the jumbled scene. Based on the average NSS scores calculated per scene category, the 12 scene categories included in our study can be placed in this order from low to high NSS scores: Jumbled < Outdoor Natural < Pattern < Cartoon < Social < Indoor < Inverted < Outdoor Man-made < Random < Fractal < Art < Action. In the comparing the above mentioned NSS scores with peak MSE, assessed over all timescales, it is evident that MSE and NSS have a significant negative correlation between them (figure 10). This could mean that MSE is indexing the top-
down information (assuming images with lesser bottom-up features might have more top-down relevance) present in scenes that is required for scene perception. Note that this contrast in MSE across scenes is reliable across almost all timescales (τ ≥ 5; figure 6B). This means that the contrast between scenes with low NSS (possibly, high top down content) and high NSS (high bottom up content) is reliable at finer as well as coarser timescales. We do not see a cross-over in MSE curves (in figure 5) as expected from our prediction about entropy differences relating to bottom-up vs top-down relevance of scenes. In other words, if individuals viewed scenes that carried more bottom-up features by making more fixations over shorter durations (faster/quicker), than when viewing scenes that had lower NSS scores (interpreted as having high top-down relevance), then absolute MSE value of scenes with high NSS scores would have been probably higher than MSE value of scenes with low NSS scores at finer timescales. Similarly, the relative MSE values of low vs high NSS scenes would have been reversed at coarser timescales – resulting in a cross-over in MSE plots somewhere between finer and coarser timescales. This is not what we see in the current study, but datasets that have/show a better separation of bottom-up vs top-down aspects of images can be used in the future to more clearly test this prediction.

In addition to the above predictions, we also tested to see if MSE will differentially index the complexity of gaze displacement signals obtained from viewing different scenes in instances where raw EM metrics, such as number of saccades and saccade amplitude were not significantly different across scene types and after removing these metrics from MSE. Our results showed that MSE indexed a significant and reliable scene contrast (figure 6) even when number of saccades did not (table 1), and even after removing (regressing out) the number of saccades from MSE (figure 7). However, saccade amplitude alone did index a scene contrast (table 1). Also, when saccade amplitude was removed from MSE, MSE still indexed a significant and reliable contrast,
but based on error bar variability, this contrast was mainly driven by only outdoor natural scenes (figure 8). Note that if we increase the BSR threshold to 3, the results are reliable over more finer timescales. This suggests that if a higher BSR threshold is used, MSE is seen to capture information beyond the information inherent in saccadic amplitude. This information captured by MSE is obtained from visual events happening closer together in time. In our subsequent post-hoc analysis, we observed a significant positive correlation between peak MSE and average saccade amplitude across scene categories, suggesting that both metrics index similar differences among scenes (figure 9). These results collectively suggest that MSE might be associated more with saccade amplitude than other basic metrics because it is sensitive to “spikes” (large saccade amplitude) in signals. This could mean that a large part of the EM signal complexity can be attributed to the saccade amplitudes in the signal, and the extent (amplitude) of our saccadic eye movements is instrumental in encoding useful information. Also, as with peak MSE, saccade amplitude (but not number of saccades or fixations) had a significant negative correlation with NSS scores across scene categories (figure 11), suggesting that both measures track with the information contained in scenes/encoded during viewing in similar ways. MSE was calculated on a continuous signal as opposed to from static regions of gaze signal data. Therefore, it is not entirely based on saccade amplitude, but instead reflects the continuous nature of eye displacement signals (encompassing discrete metrics such as saccadic number, amplitude etc.), and can be treated as a distinct complexity metric. Also, in theory, it is a direct measure of the information encoded by eye movements during viewing and contained in gaze signals. However, based on analyses involving saccade amplitude, MSE and NSS scores, it seems that saccade amplitude might be a more conventional and easy way to index the same scene effects that MSE is capable of indexing.
This study suggests that MSE can be used as an index of the complexity of gaze signals as they unfold over time, and as an extension, the information encoded by these signals during viewing. Hence, it contributes to the effort made by researchers to devise new metrics of EM signal complexity. However, it is important to note that this is a single study that has explored the possibility of MSE serving as an eye movement complexity metric. Future studies that must be conducted to validate the use of MSE as an eye movement complexity metric include, applying this metric on gaze signal data collected from different groups of individuals (e.g. older vs younger adults, amnesic patients etc.) and different experimental paradigms (e.g. face versus scene viewing). Moreover, future experiments can also investigate if MSE, as already established with other EM metrics (e.g. number of fixations), is associated with cognitive processes such as memory and other neural signals (e.g. neural MSE), in order to fully identify its usefulness as a EM based marker of neural processing.

3.5 Conclusion

In summary, in this study we have shown that Multi-scale entropy (MSE) can be used to empirically measure the degree of complexity of gaze signals that follow $1/f$ power laws (correlated signal events are present over longer time intervals) produced from viewing scenes over a defined period of time. Since entropy relates to information, we suggest that MSE also measures the information encoded by these signals during viewing. MSE differs between scene types suggesting that it can index differences in the information encoded from different types of scenes. Based on saliency-based scene classification done in Borji & Itti (2015) and the contrast in scenes indexed by MSE in our study, one possible conclusion is that MSE might be contrasting scenes that have more bottom-up features vs top-down relevance – higher MSE might be related to higher top-down information encoded in gaze signals for scene perception.
Finally, MSE also establishes differences between scene types even in instances when certain basic EM measures (e.g. number of saccades, fixations) do not, and even when these measures are removed from MSE. However, we do note that of the basic EM measures considered in this study, MSE is positively associated with saccade amplitudes, suggesting that complexity of EM signals and information encoded in these signals might relate with the extent of saccadic eye movements that we make. We also see that saccadic amplitude can track scene differences similar to MSE and can therefore serve as a more conventional and easy measure that is derived from raw EM data. Since MSE is based on the continuous EM signal data and is an established marker of information contained in signals, it could still be treated as a distinct measure of EM complexity.
Chapter 4

4 General Discussion

The purpose of the current thesis is to introduce a metric that could empirically quantify the changes in eye movements, i.e. eye movement dynamics, during viewing. The metric that we have applied for this purpose is Multi-scale entropy (MSE) which has been previously applied in research studies to quantify the dynamics of various types of signals (Costa, Goldberger, & Peng, 2005; Heisz, Vakorin, Ross, Levine, & McIntosh, 2013; McIntosh et al., 2013). This thesis is built on two main hypotheses based on our review of the literature on eye movements and signal dynamics. The first hypothesis is that if eye movements recorded during viewing can produce a signal that shows complexity, which in turn makes this form of signal informative, MSE can be applied on the signal to quantify the complexity and information contained therein as viewing unfolds over time. The second hypothesis is that MSE can differentially quantify how the signal complexity (and information) of gaze signals changes with the type of stimuli being viewed – suggesting that it could be used as a measure of eye movement variability that is based on the type of stimuli.

The first hypothesis was devised on the basis of previous studies done in the field of vision research and applications of MSE. It is known from previous studies that raw eye movement recordings can be converted to gaze signals on which various techniques can then be applied. One such application involves converting eye movement patterns recorded during viewing into gaze displacement (positional difference between gaze fixations) signals on which signal processing methods such fast fourier transform and autocorrelation procedures can be performed to characterize the dynamics of these signals (Aks, 2008; Aks, Zelinsky, & Sprott, 2002).
Accordingly, when these techniques were applied on gaze displacement signals obtained from previous viewing studies, researchers noticed patterns in signals (Aks, 2008; Aks et al., 2002; W-L Wu, Anderson, Bischof, & Kingstone, 2014; Fairbanks & Taylor, 2011). For example, studies have shown that gaze fixations are made repeatedly on a given image position over close as well as distant time intervals, suggesting that eye movements produce these recurring signal patterns over many different time intervals as viewing unfolds over time (W-L Wu, Anderson, Bischof, & Kingstone, 2014; Aks, 2011). Since eye movements are not restricted to any one time interval, and constantly change over the time course of viewing, they are considered to be temporally dynamic entities. Similarly, different types of eye movements, including saccades vs micro-saccades operate over different spatial scales of resolution (as measured in pixels on the screen), showing that eye movements are also spatially dynamic (Aks, 2008; Aks, Zelinsky, & Sprott, 2002; Fairbanks & Taylor, 2011). Finally, these spatial differences in eye movements also tend to recur over the course of viewing, indicating that eye movements can simultaneously show spatial and temporal dynamics/changes during viewing and produce complex signals (Aks, 2008; Aks et al., 2002; W-L Wu et al., 2014; Aks, 2011; Fairbanks & Taylor, 2011). However, as one would imagine, the nature/type of complexity of these signal patterns changes with the type of visual task individuals are asked to perform (e.g. different types of visual search tasks) and types of signal on which NDS techniques are applied (e.g. absolute gaze fixations vs gaze displacement; Fairbanks & Taylor, 2011).

Of the various types of signal patterns that eye movements are known to produce (as documented in previous studies), one particular type of signal that is of importance to the first hypothesis of the current thesis is the $1/f$ type of signal. The signal dynamics literature characterizes $1/f$ as a type of (colored) signal (similar to other colored signals like brown and black) that consists of structures that correlate with each other over longer time intervals compared to white noise.
(which has no correlated structures at this time interval). In the case of gaze signals, fixations that are made repeatedly over longer time intervals on a given image position is an indication that the signal is of the type $1/\alpha$. In vision research, this type of gaze signal is considered to be meaningful because it suggests that viewing is not an entirely random process, but that contingencies might exist across gaze fixations produced during viewing (Aks et al., 2002). Aks et al. (2002) goes as far as to say that this type of viewing could be used to optimally encode information from a complex environment, which is then laid down in the form of a long term spatial memory across these fixations.

In the signal dynamics literature, the complexity of $1/\alpha$ type signals has been quantified by MSE (Costa, Goldberger, & Peng, 2002; Costa et al., 2005). Costa et al. (2005) showed that MSE is capable of assigning high entropy (signal irregularity) values to $1/\alpha$ signals compared to both white noise that is relatively meaningless and completely periodic signals. The difference in entropy as indexed by MSE is especially pronounced at coarser timescales, which correspond to signals captured over longer time intervals. Furthermore, based on the idea that entropy is equal to the concept of information (Shannon, 1948), Costa suggested that MSE reveals that $1/\alpha$ type signals contain high entropy and hence, high information content (hence, more meaning) compared to white noise or periodic signals. To sum up our first hypothesis, we expected MSE to quantify both the temporal (entropy across different timescales) and spatial (information encoded from the visual field) dynamics of eye movements, provided that the gaze signals produced by eye movements belonged to the $1/\alpha$ class of signals.

The second hypothesis relates to the applicability of MSE as a metric that can quantify differences in eye movement patterns produced from viewing different types of stimuli. We know from the vision literature that characteristics of stimuli (e.g. scenes, faces etc.) guide
viewing in specific ways, including influencing the number of fixations we make on the image, duration over which we look at images etc. (Antes, 1974; Henderson, Weeks, Phillip, & Hollingworth, 1999; Loftus & Mackworth, 1978). One aspect of image viewing that helped form the 2nd hypothesis of the current thesis is the viewing differences arising from differences in the bottom-up vs top-down cues present in images. For instance, research shows that distinct bottom-up features of images such as unpredictable contours elicit more fixations during viewing compared to smooth surfaces (Mackworth & Morandi, 1967). With respect to top-down relevance, it is known that components of images that are semantically out of place in an image would again elicit a higher number of fixations than parts of images that one would expect to see in the image (Loftus & Mackworth, 1978). Research done by Henderson et al. (1999) shows that over the course of viewing, bottom-up features dictate viewing faster/earlier than semantic/top-down cues. Therefore, keeping these studies in mind, we hypothesized that MSE, if shown to be an indicator of the complexity and information contained in gaze signals (hypothesis #1), should differentially assign entropy values as a function of timescales to different types of images shown to participants. In this way, we hypothesized that it could serve as a measure of eye movement variability that is driven by the type of stimuli.

In order to address these hypotheses in the current thesis, we carried out a study in which individuals were asked to perform a free viewing task in which they viewed images taken from several categories, including art, natural and artificial scenes etc. (images were taken from: Borji & Itti, 2015). The eye movement recordings were then processed into gaze signals and MSE was applied on these signals. Our interpretations of the results from this study seem to support our two hypotheses. Results showed that our gaze signals followed 1/f signal dynamics, supporting (based on previous research) the use of MSE on these signals. When MSE was applied on gaze signals, it assigned high entropy values to finer and coarser timescales (except very fine
timescales), thus quantifying the high complexity and information encoded by eye movements (during viewing) and contained in these gaze signals. Next, MSE also significantly differed across scene types, indicating that it can index eye movement variability that is driven by the type of images. The image categories used in the current study were assigned average normalized scan path saliency (NSS) scores, computed from the NSS scores per image category, obtained from the Borji & Itti (2015) dataset. Based on the fact that these NSS scores were an average of scores produced from three saliency models, all of which used bottom-up image features to assess saliency, one possible interpretation of NSS based classification is that images with high NSS scores contain largely bottom-up cues, whereas images with low NSS scores have more top-down relevance (rather than bottom-up relevance). Incidentally, in our study, images that were assigned high entropy values by MSE also turned out to be images that had low NSS scores (MSE was negatively related to NSS). Based on our interpretation of NSS score-based image classification (low NSS = high top-down relevance; high NSS = high bottom-up relevance), it might be the case that MSE, through assigning high entropy values, is indexing the cognitive information needed for scene perception. However, since we do not relate MSE to any cognitive outcome (e.g. time for perception, recognition etc.) in this study, this inference is not definitive. Finally, MSE assigned high entropy values at all timescales across all images in our study, suggesting that visual events (e.g. saccades) that occur at close as well as distant time intervals add complexity to gaze signal patterns, and are useful in gathering information that is useful to distinguish between scenes. Hence, so far, our results suggest that MSE can serve as a metric of gaze complexity and information encoded through eye movements, and that it can also index stimulus-driven variability in eye movements.

As an extension to this study, we also decided to investigate how MSE relates to basic measures of eye movements that are already used extensively in vision research. These basic measures are
conventionally thought to be static measures as they provide a one-dimensional view of vision, e.g. by quantifying the number of fixations made on a given position or the duration of viewing etc. This idea can be contrasted with MSE, which captures gaze complexity across different time windows (timescales), thus capturing how gaze complexity changes as viewing unfolds over time – thus quantifying the dynamic aspect of viewing (real-time viewing). In our study, we found out that of the three static measures of eye movements we analyzed (number of fixations, number of saccades and saccadic amplitude), only the saccadic amplitude differed significantly across scenes. Furthermore, in regressing these static measures individually from MSE, we found out that the resulting residuals still showed a significant scene contrast in the absence of these static measures, suggesting that MSE can be treated as a distinct metric of gaze complexity. However, it is important to note that when we analyzed the relationship between saccade amplitude and MSE, there was a significant positive correlation across scene types. Since MSE as a metric is influenced by ‘spikes’ in data (which resembles saccadic jumps in gaze signal data), this association, is to a certain extent, a result that is expected. Saccade amplitude is also seen to track with NSS scores similar to MSE. Therefore, it is possible that MSE is mostly capturing the information inherent in the saccadic amplitude. However, since MSE is applied on a continuous stream of signal data that encompasses fixations, saccades etc., it could still be treated as a metric that captures the information encoded by the signals in totality, as these signals (produced by eye movements) unfold over time.

Although the study that has been documented in this thesis addressed our two hypotheses to a certain extent, we do recognize that there are gaps that need to be filled with future work to completely address our broader aim of developing MSE as a marker of eye movement variability. First, our hypotheses are based on the application of MSE on gaze signals that follow a particular type of complex signal, i.e. $1/f$ signal. However, as mentioned earlier, the type of
signal dynamics changes with the type of visual task performed and gaze signal analyzed. For instance, previous work has documented instances where gaze fixation signals and gaze displacement signals at higher signal frequencies (compared to $1/f$) showed Brownian behavior, characterized by highly correlated signal structures over time intervals that are relatively shorter than the time intervals for $1/f$ signals (Aks et al., 2002). These results suggest that gaze signals can show different types of colored signal behavior (specific type differs by the time durations over which one sees correlated signal structures) as opposed to white noise (no correlated structures at longer time intervals). Therefore, it is important to apply MSE on other types of gaze signals (e.g. absolute fixation data, direction of fixations etc.) that show different colored signal dynamics (e.g. Brownian, black power laws) at different frequencies, in order to fully understand the utility of MSE as a general (and not $1/f$ specific) marker of eye movement (gaze) complexity.

Next, the concept of timescale as defined in the literature on MSE has been adapted to specifically suit gaze signals produced in the current study. Timescales relate to the time between successive visual events, and in the current study, timescales are defined as coarser scales if the saccadic events are seen to decrease in number based on a preliminary observation of the signal. In the current study, this happened at around a timescale of 60. Hence, all timescales below 60 were clustered in the finer timescale category. However, it might be important to construct a more rigorous definition of coarser vs finer timescales across trials such that it applies to all or most gaze signals produced from eye movements recorded at the sampling rate used in the current study (i.e. 500 Hz). It might also be useful to create a system to delineate timescales based on the sampling rate of the instrument so that one can use eye movement data collected at a different sampling rate for MSE analyses in the future.
Moreover, in addressing our 2nd hypothesis, we proposed one possible explanation for the scene contrast indexed by MSE. Our results showed that MSE assigned high entropy values overall to images with low NSS scores and relatively low entropy values to images with high NSS scores. Since NSS was computed in Borji et al (2015) from bottom-up features, we suggested that low NSS could indicate that the image has more top-down relevance and high NSS could point to more bottom-up relevance. Therefore, we interpreted that MSE, by assigning high entropy values to low NSS images, could be indexing the information necessary for perceptual processing and semantic comprehension of the scenes. However, this is only one way of interpreting the data and it is possible that the relationship between high entropy and low NSS values that we see in our results is due to a third variable that correlates with NSS. Also, from an analytical perspective, it might be more powerful to relate MSE to NSS scores per image directly to probe into the association between the two further, than using categorical scene classification as in the current study. The reason for using such classification in the current study was to also look if MSE can index a scene contrast on its own, as well as after removing the effects of basic metrics (number of fixations, saccades and saccade amplitude) from MSE. In other words, PLS, which was used to statistically assess these contrasts needs study condition (e.g. in this case, scene category) as an input.

When thinking about the broader implications of MSE as a marker of eye movement complexity, it becomes important to identify the usefulness of this metric as a marker of cognitive processes as well. This is important because eye movements have been associated with higher level cognitive processes, and most previous studies suggest that eye movements both reflect and inform such processes, especially memory (Heisz & Ryan, 2011; Olsen et al., 2015, 2016; Ryan, Althoff, Whitlow, & Cohen, 2000). While we have suggested in the current study that MSE when applied on gaze signals quantifies the complexity and associated semantic information
required for the perception and/or comprehension of scenes, it is important to directly correlate MSE with other cognitive measures like recognition memory scores, to specify its usefulness as an eye movement-based index of cognition. It is also important to carry out such studies across different individuals, including amnesic individuals, older and younger adults, as it has already been done with conventional eye movement measures (Chan, Kamino, Binns, & Ryan, 2011; Firestone, Turk-Browne, & Ryan, 2007; Olsen et al., 2015, 2016; Ryan et al., 2000). In particular, MSE has to be applied on eye movement data collected from studies that contrast conditions (e.g. orientation of face stimuli: Olsen et al., 2015) and age groups (older vs younger adults: Firestone et al., 2007; Memory impaired individual vs healthy controls: Olsen et al., 2016) in relation to a particular cognitive dimension (e.g. recognition memory). Also, we could investigate if MSE can identify differences in gaze patterns elicited by viewing stimuli that probe different memories (e.g. autobiographical vs semantic: Heisz et al., 2013).

While we understand these gaps in the study, we could not rectify some of these gaps and broaden the scope of this thesis due to constraints in time and study design. Specifically, the study used in this thesis was designed to be a free scene viewing task that did not have a cognitive component associated with the task. Hence, we had no means of relating the MSE of gaze signals produced from eye movements to results of a cognitive task to identify whether MSE of eye movements can predict higher level cognitive processes. Moreover, with respect to stimulus classification, since NSS values were only available for each category of images as opposed to each individual image, we could not quantitatively regress these values from MSE values for each image to then numerically document the effect of NSS scores on MSE. For example, if this analysis was possible, we could have identified whether MSE loses its ability to index the scene contrast after removing NSS scores from it, in which case we can more
concretely say that MSE is based on NSS differences across scenes, and that it predicts the top-down/cognitive information needed for scene perception.

In essence, we consider this thesis to be work that contributes to the growing literature on the quantification of eye movement dynamics by Aks et al (2002, 2005, 2008) and others, and lays the foundation for further verification of MSE as a marker of eye movement complexity. In this novel project, we have applied MSE, a tool that has been used in other fields as an index of signal complexity, on gaze signals produced by eye movements recorded when individuals freely viewed various types of scenes. From our results, we suggest that MSE can be treated as a distinct marker that quantifies the complexity of gaze signals, or in other words, the information contained therein, as viewing unfolds over time. This information encoded during viewing (and contained in gaze signals) cannot otherwise be numerically quantified in real time. Based on study specific results, our interpretation is that MSE might also be indexing the degree to which cognitive/top-down information gained from scenes dictate ongoing viewing, providing a reason to further investigate the direct association between eye movement based MSE and higher level cognitive processes, including memory.
Chapter 5

5 Conclusions

In summary, we have applied multi-scale entropy (MSE), a metric that has been used to quantify the complexity of signals, on gaze signals produced from eye movements. In general, results from our study suggest that MSE assigns high entropy values at all timescales to gaze signals of the type \(1/f\), suggesting that it can successfully quantify the complexity (visual information encoded) and dynamics of gaze signals produced by eye movements. Next, MSE also distinguished between scene types in our study, suggesting that it could index complexity in eye movements that is driven by the type or information content of stimuli. Specifically, our interpretation of the results is that it indexes the cognitive information needed for scene perception and/or comprehension based on the fact that it assigned high entropy values to images that seemingly have low bottom-up feature scores. Finally, we also suggest that MSE can serve as a marker that is distinct from the static measures of eye movements that are conventionally used in the field of vision research because it captures eye movement complexity across scene types in instances where certain measures like number of fixations/saccades did not significantly differ across scenes. Our hope is that this work would lay the foundation for further research that focuses on verifying the usage of MSE as a marker of eye movement variability and more importantly an eye movement-based marker of cognitive processes, especially memory, across different groups of individuals.
Chapter 6

6 Future Directions

As the current thesis is only the first step towards investigating the implications of using MSE as a metric of eye movement variability, more work can be done in this area to definitively verify its usage. Specifically, future work can be clustered into three different types: extending the current scene viewing study, devising alternative study designs to explore related research questions, and performing comparative analyses with results obtained from the application of MSE in other related modalities, e.g. neural MSE, to understand the implications of MSE as an eye movement variability measure, both in vision research, and in the long run, in clinical populations.

First, the current study can be extended by testing different groups of individuals in the scene viewing study used in the current thesis. This would give us variable eye movement recordings from which gaze signals can then be computed for the application of MSE. This would then inform us of whether and how MSE differentially assigns entropy across timescales, across the different individuals. We could also compare differences in eye movement based MSE with differences in conventional static metrics such as number of fixations/saccades, saccade amplitude etc., as documented in previous studies, across individuals. In particular, testing memory impaired individuals versus health controls can be used to identify if MSE differs between these two groups. Such differences would then indicate the kind of information captured by MSE, that is useful in supporting memory processes in controls vs memory impaired individuals.
Second, it might also be useful to devise new study designs to obtain a more extensive eye movement data set to apply MSE on. For instance, we currently have certain eye movement-based phenomena in the literature, including the eye movement repetition effect and relational manipulation effect. Repetition effect refers to the idea that the number of fixations made on an image would decrease with the number of times an individual is exposed to the given image. Therefore, it would be interesting to see if MSE of eye movements correlates with the repetition effect shown by the static fixation count measure. This would mean that MSE quantifies the information required to develop a memory representation for an item. Also, relational manipulation effect involves increased viewing in areas of scenes where the relationship between objects has been altered. If MSE correlates with viewing to altered scenes, this could mean that it is capturing the information needed to lay down spatial memories of items. Both scenarios would also suggest that MSE, while serving as a relatively ‘dynamic’ metric still retains the qualities of conventional metrics and can be compared with current studies for validation/other research purposes.

Moreover, many studies show that certain eye movement measures are directly associated with cognitive processes, particularly, memory (Chan et al., 2011; Olsen et al., 2016). Hence, in order to find out if MSE of eye movements is also related to memory processes, we could design a study that contains a memory task associated with it, that would help us gather scores (e.g. recognition memory ratings), which can be directly compared with MSE. This would serve as a direct way of identifying what type(s) of memory processes are indexed by MSE. The same concept can be extended to other cognitive domains, such as attention, perception etc. and tested on various groups of individuals including older and younger adults who are healthy and also those who show cognitive impairments, to again help us identify if these cognitive operations could be reliably indexed by MSE.
Finally, in terms of broad implications, one concept that was discussed during the development of this thesis is the possibility that eye-movement based MSE could serve as a proxy to neural MSE. MSE has been used already on neural signals to quantify the complexity of brain signals across certain groups of individuals and study tasks (Heisz et al., 2013; McIntosh et al., 2013). Similarly, studies have also shown direct structural connections between the oculo-motor (vision) and hippocampal (memory) systems of the brain (Shen, Bezgin, Selvam, McIntosh, & Ryan, 2016), as well as a functional relationship between hippocampal activation and eye fixation measures (Liu, Shen, Olsen, & Ryan, 2017). Building on this work, it might be useful to compare MSE of eye movements with neural MSE, to see if the structural and functional relationships established between the visual and memory systems also extends to the complexity of signals that are individually produced by both systems. This would also help quantify the relevance of visual information encoded during viewing to neural information processed in the brain and as well as the potential inverse association between the two types of information. If established to be similar or related to one another, MSE that is based on eye movements can serve as a proxy to neural MSE.
References


https://doi.org/10.3758/BF03210264

https://doi.org/10.1068/p261059

https://doi.org/10.1163/156856896X00123


https://doi.org/10.1523/JNEUROSCI.3987-14.2015

https://doi.org/10.1016/j.cortex.2016.10.007


https://doi.org/10.1167/9.10.2.Introduction


https://doi.org/10.1007/s004260100059


SR Research Ltd. (n.d.). Mississauga, ON, Canada: SR Research Ltd


