TIME SERIES ANALYSES OF HOLLYWOOD FILM AND REACTION TIMES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jordan Edmund DeLong

January, 2015

TIME SERIES ANALYSES OF HOLLYWOOD FILM AND REACTION TIMES

Jordan Edmund DeLong, Ph. D.

Cornell University 2015

The aim of this dissertation is to explore the link between long-term attention and memory by looking at two types of data, Hollywood film and reaction time data. Data from a variety of sources are analyzed using a sequence of techniques designed to detect patterns in the data related to dimensionality. These patterns have been linked in to long-scale fluctuations in human attention, amongst other natural phenomena, but there is much research to be done before concluding whether long-term temporal structure plays a meaningful role in human attention. This question is investigated through four main chapters: (1) An introduction to the concept of timeseries analysis, fractal structure, and how these phenomena may be linked to neural systems; (2) An analysis of the structure of shot lengths in Hollywood film, with several techniques simulated and compared; (3) A review and analysis of historical reaction time data collected before the advent of film; (4) A series of novel laboratory experiments which investigate the link between attention and memory for visual stimuli presented in different temporal structure.

BIOGRAPHICAL SKETCH

Jordan E. DeLong earned Bachelor of Science degrees in 2007 from Indiana University, majoring in Psychology and Cognitive Science. While at IU, he worked in the lab of Tom Busey, earning the biannual Outstanding Research Award and the J. R. Kantor Prize, given annually for the best undergraduate honors thesis in psychology.

Dr. DeLong joined the doctoral program in Psychology at Cornell University, with an emphasis in Perception, Cognition, and Development while working in the labs of James Cutting and David Field. During his years at Cornell he received two years and four summers of research support from the Sage Fellowship as well as completing eleven publications in peer-reviewed journals and invited chapters. He proposed and taught multiple courses, including a course introducing the study of the brain through network theory and complex systems, courses teaching writing for scientists, and a programming course entitled “MATLAB for the Behavioral Sciences”, which lead to a Graduate Student Teaching Award in 2010. During the summer of 2013, he taught a course on learning techniques to first-semester student-inmates at Auburn Maximum Security Prison as part of the Cornell Prison Education Program.

After leaving Cornell, Dr. DeLong returned to Indiana University as Visiting Assistant Professor. He is the sponsor of a student group which specializes in helping local businesses and nonprofits apply psychological concepts to their business decisions and mentors students underrepresented in STEM fields.

To Helen Nita, who believed I could do anything.

ACKNOWLEDGMENTS

My academic life and the work presented in this dissertation would never have been possible without a number of people who have helped me become a better person and academic throughout the years.

First I need to thank my love, partner, and fiancée Devan, who makes my life better and more interesting every day. We made it through this together and better than ever.

I thank my family, including Mom, Dad, and Sarah, who have given me so many gifts and have made me who I am today. They are so easy to take for granted. And to my best friend Kevin, who is the closest thing to a brother I will ever know.

To my friends from Indiana, including Tom Busey, Bethany Jurs, Dean Wyatte, Alex Burkhardt, and Melissa Troyer. I think I will always be reaching to recreate that first lab experience I had with all of you.

I am immensely proud to say I have learned most of what I know from James Cutting and David Field. I feel blessed that I got to spend years observing and emulating two men who are both brilliant scholars and wonderful people. Watching Vivian Zayas go from new hire to tenured faculty was inspiring, and made be believe that good people can win. Listening to Barb Finlay makes you remember that while anyone can shout and argue, the most important statements are said quietly, only revealing their impact after taking root and slowly growing inside your mind.

There are also so many people who made me appreciate my graduate school experience: my lab-siblings Dan, Mark, James the Younger, Kedar, Kate, Catalina, and Ayse; undergraduate lab members like Christie Nothelfer, Lucy Zheng, Alexandria Smith, Lizzie Brooks, Peter Cohn and Shayna Ratner; staff who have bailed me out on multiple occasions like Pam, Cindy, Linda, Mary-Lou, Julie and Fred; and the other faculty and graduate students who make up the weird, wonderful, and brilliant cast of characters at Uris Hall.

Finally, I want to thank Debra Krueger, my second grade teacher, who made me love being a student. I have had many great teachers throughout the years, but the path to completing this dissertation started in her classroom at Harris Elementary.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

A 1/*f* primer

Frequency and Power

In the interest of the reader, I would like to begin by informally framing a single concept that will be referred to relentlessly throughout every major section of this dissertation. While the following sections of this dissertation are interrelated and sequential, they are also designed to be mostly self-contained. Rather than repeating or presenting a stripped-down description three times, this opening section will attempt to present a singular, satisfying, and accessible introduction to the concept of the 1/*f* pattern and why it is of interest to psychology.

Mathematically, a 1/*f* pattern is simply a signal that has a particular relationship contained within the strength of different frequency components. These components can be easily computed utilizing Fourier analysis, a technique that recodes any signal using sine and cosine waves. Investigating the relationships between wave components of the signal yields useful information about the composition of a dataset. While continuous dataset may appear mostly random to the eye, Fourier analysis can reveal regular oscillations at different frequencies that may be obscured due to the noisy nature of most natural signals. Because frequency is defined as a change over time, these measures are *always* taken over a number of samples, with the number of samples per unit limiting which frequencies can be computed.

Looking at the relationships between frequency components in statistical programs or packages such as MATLAB (2013) or R (2014) is very similar to looking at the equalizer output on iTunes or a hi-fi stereo system. Instead of looking directly at the fast-moving voltages that drive a vibrating speaker, audiophiles can observe how much each frequency band is contributing to the overall volume by viewing a binned histogram of ‘power’. Songs do not contain equal volume in each frequency bin. In general, rap songs have more volume distributed within lower frequencies (bass) than barbershop quartets. What is driving this discrepancy? Obviously, Dr. Dre uses a range of musical instruments and devices that create epic bass beats while the range of tones that an a cappella group can muster is more restricted. This does not mean that Dre will <u>always</u> utilize more low frequencies for every second of the album, but will exhibit greater power in the lower frequencies with a long enough sample. Users reset the EQ on their iPhones without thinking twice about the sophisticated filtering and analysis going on behind the scenes.

Using Fourier analysis to compare the contribution of different frequencies is not just a way to compare the differences between genres, but is powerful enough to tell the difference between individual songs. Music identification services like Shazam compute a histogram of frequency/power information from ten seconds of audio recorded through a cellphone (Wang, 2003, 2006). This ten-second sample contains enough unique information to distinguish an individual track from over 11 million other songs given a clean enough recording. It is not the case that the histogram version of the song clip contains *more* information when it is in the frequency domain, but it is simply a *different* way to view the characteristics of the song. In this case, categorization is made much simpler when translating from the time domain (waveform) to the frequency domain (frequency and power).

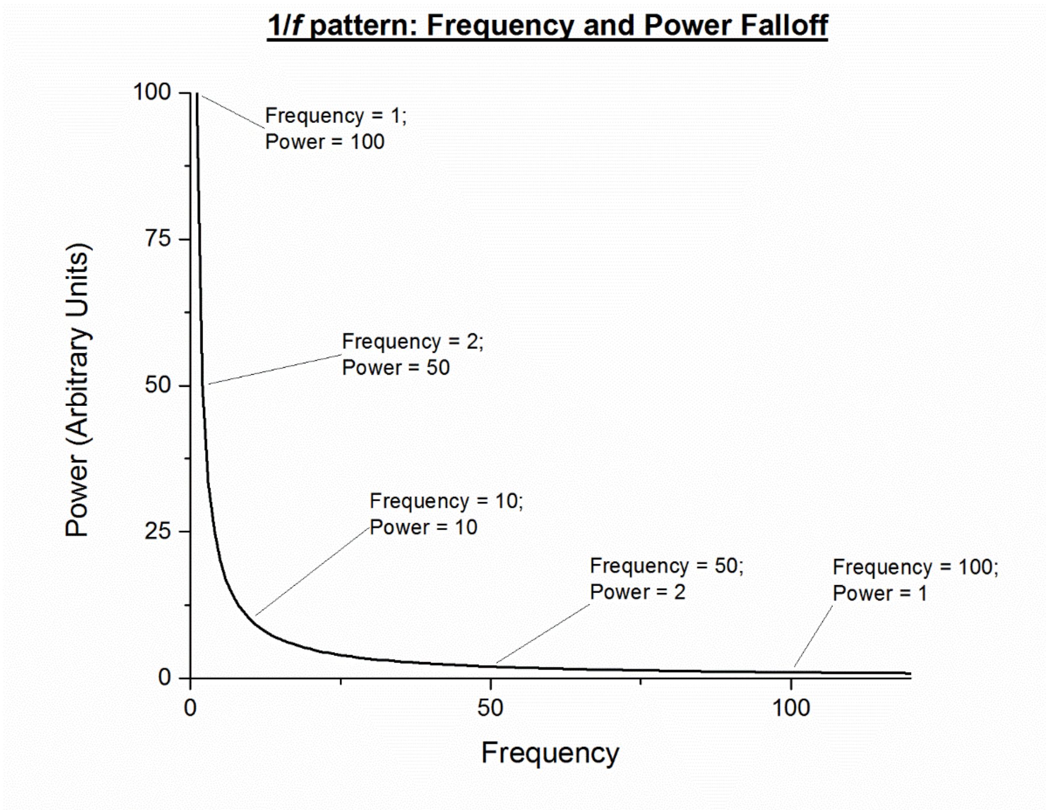

Figure 1.1 Power falls off exponentially with frequency in this idealized 1/f distribution

The technique of comparing the amount of energy contributed to different frequences is not confined to music. As hypothesized by Fourier, *any* discrete signal can be mathematically decomposed and recomposed using sine/cosine waves, which are easy to combine and decompose. Engineers take entire courses about the technique and utility of the one-dimensional Fourier analysis (Osgood, 2007). Research in vision science has looked at the spectral components of different visual stimuli, utilizing two-dimensional Fourier analysis to analyze the statistics of natural scenes (Field, 1987) and artwork (Graham & Field, 2008). Another type of dataset is an ordinal time series, in which events are not compared using signal amplitude but rather the amount of time elapses between events. Regardless of the type of data, the fundamental technique of Fourier analysis is the same – only the interpretation changes.

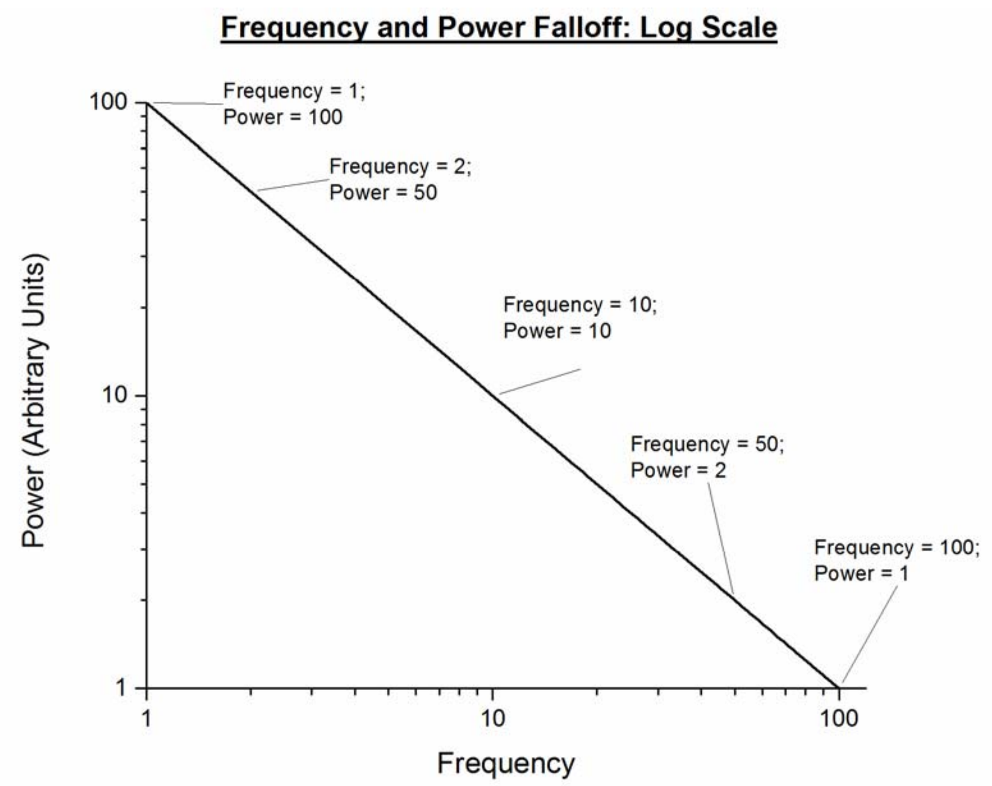

Figure 1.2: The 1/f pattern appears as a straight line on a plot with logarithmically scaled axes.

How to define and identify the 1/f pattern

The type of relationship between power and frequency that constitutes a 1/*f* pattern is explicitly stated within the name – the power falls off at a rate of 1/*frequency*. This means that with a 1/*f* pattern you could observe a power unit of 100 at a frequency of 1, 50 units at a frequency of 2, 10 units at a frequency of 10, 1 unit at a frequency of 100, and so forth as seen in Figure 1.1. This specific pattern is a type of power law, a class of functions modeled using exponential functions.

The attribute that makes a 1/*f* pattern notable and distinguishable from other power laws is that the 1/*f* pattern has the same relationship between its components at every scale. This means that, if you look at the relationship between large and small components at a wide scale, the curve of the graph looks exactly like a zoomed in

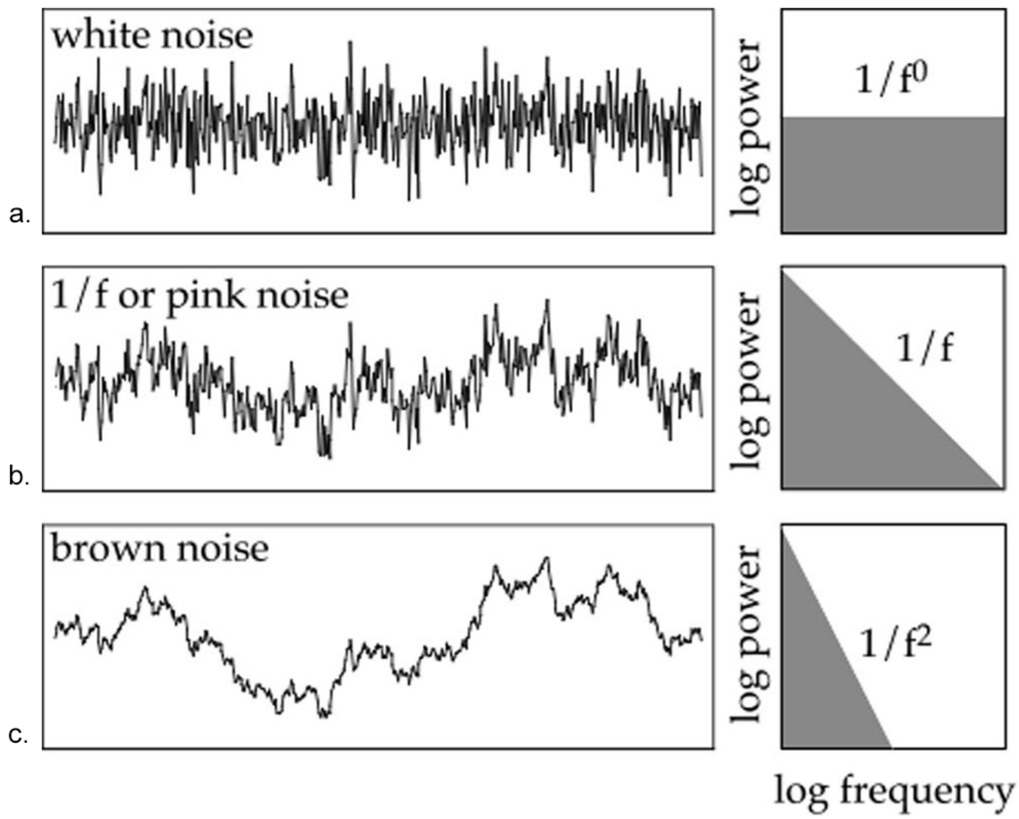

Figure 1.3 The visual approximation of three types of noise referenced in this paper: white (random) noise, 1/f or pink noise, and the random walk or brown noise. The figures to the right represent the power vs. frequency falloff on a log-log plot. Figure is from (Gilden, 2001).

version. In fact, *any* part of an idealized 1/*f* distribution looks *exactly* like the curve of any other part of the distribution.

Most graphs of 1/*f* patterns are put on a logarithmic scale for the simple reason that viewing them on a fixed linear scale causes the majority of the data to hug both axes of the graph. The logarithmic scale also displays that when the data from a 1/*f* source is plotted on a log-log graph (a graph where both axes are logarithmic) the data typically form a straight line down the middle of the graph (as seen in Figure 1.2). The most straightforward way to test whether something is 1/*f* is to take the Fourier transform of the data and check the slope of the line on a log-log plot. Because a 1/*f* distribution is a straight line at each scale, in order to be categorized correctly as 1/*f*, it has to have a slope of -1. This same test can identify two other kinds of noise: white noise (truly random) and Brownian noise (highly ordered noise). This method can be easily modeled using a linear fit and is resistant to noise.

White noise is wholly unpredictable from one sample to the next (Figure 1.3a). To extend the audio analogy, white noise is what we get when we turn on a radio that is not tuned to any station. The current value the signal does not influence the next point in the signal, so each frequency is equally likely. This means that the graph of power and frequency will present as a straight line as any and all changes are equally likely over a given time period.

Brownian noise is also known as a ‘random’ or ‘drunkard’s’ walk (Figure 1.3c). While this noise is still random, it is random in a much more structured way. Imagine a person flipping a coin to decide whether the next value in a dataset should be higher or lower than the previous value. After adjusting higher or lower, the person flips again to get the next value. Anchoring where the next value to the previous value means that unlike white noise, the previous value in the signal plays a major role in determining what the next value will be. Brown noise is actually named after Brownian motion, the type of motion observed when you put a tiny speck of pollen into a jar of oil. The speck can move any direction within the jar, making its final position practically unpredictable, even if the path it travels is predictable in the short term. Brown noise shows up as a slope of -2 on the log-log plot for a one-dimensional timeseries like the ones used in this dissertation.

$1/f$ noise presents as a slope of -1 on a log-log plot, exactly between the slopes for white noise (true random) and brown noise (random walk). This placement is due to the fact that the previous point in time may influence the next point; the possible range of variation is constrained more than white noise but less than brown noise. The sweet spot directly between these two types of noise is 1/*f*.

Why care about 1/f?

These characteristics, however, are not enough to explain *why* academics notice and comment on 1/*f* patterns. The easy answer to why 1/*f* patterns are remarkable is that they are found in a variety of locations with no discernable explanation for their presence. Of course, many distributions are considered standard (like the bell curve) but those typically appear to arise for fairly well defined reasons. A 1/*f* distribution is slightly more exotic, and has been tied to not-yet-understood phenomena that arise in sufficiently complex systems. 1/*f* patterns have been tied to things as mundane as the emergence of electrical noise and as grandiose as the emergence of life on earth.

What defines a complex system? It is easier to explain the definition of a complex system in terms of what it is *not*. A complex system *does not* follow the simple rules that would allow them to be characterized by linear processes. However, that does not mean that the behavior of complex systems are *completely* unpredictable. While some complex systems are wholly random, some lie between pure randomness and pure order, and like unhappy families, they can be complex in different and interesting ways. Current work in complexity theory attempts to characterize these different types of processes and understand what makes them different.

A canonical definition of a complex system is still under debate, but some basic guidelines (Holland, 2006) can be paraphrased as a system where:

1. The system has many parts that behave independently.
2. Those parts influence one another's behavior.
3. The system exhibits memory, meaning that individual parts are influenced by feedback from previous states.

Other guidelines include a lack of central control, robustness to degradation, and unexpected properties that emerge from the system. While these definitions are currently workable, they will no doubt need to be changed and revised as we learn more about what happens inside complex systems as they interact. Complex systems are fascinating and frustrating because they are resistant to description using flow charts, hierarchies, and traditional modeling tools.

A currently undefined class of complex systems appear to emit 1/*f* noise as a signature. This is a rather broad statement; as 1/*f* noise exists in a number of phenomena that do not appear have any obvious connections. A 1/f pattern shows up when you observe the height of the Nile River over time, the size of asteroid impact craters on the moon, the amount of traffic on single street in Los Angeles, and even the visual structure of natural world. While it is clear that asteroids and traffic jams probably are not being driven by the same mechanism, is there something about both systems that lead them to similar behaviors?

Fractional dimension, fractals, and the brain

One major clue appears to be found in the mathematical exploration of fractals. Much of the work on complex systems began when Benoit Mandelbrot asked a simple question, “How Long is the Coastline of England?” (1967). On the surface, measuring distance appears to be a simple task– just measure along the edge of the coastline and you have your distance. In practice, the coastline of England is not a straight line that can be easily measured. Coastlines have bays and peninsulas, rivers, inlets, coves, and at the smallest scale, individual pebbles, rocks, and sand. If you measured the coastline by the mile, you would be skipping over a lot of detail than if you measured it with a yardstick. It turns out that while measuring something as simple as a coastline, the distance you observe is dependent upon what you use to measure. As you get to smaller and smaller scales, the measurement of the coastline becomes longer and longer.

The rate in which the distance of the coastline increases is directly due to the degree to which the coastline is convoluted. It is possible to find white-sand beaches that actually are somewhat straight, and the distance of the coast does not increase very much until measuring at a very fine scale. On the other extreme, coastlines like the Fjords of Norway are highly, highly convoluted waterways. It is clear that these coastlines are very different, but we need more than raw distance to describe them properly. The answer is to come up with a metric to describe *how* convoluted a coastline is, and several techniques have been developed to calculate a measure called fractional dimension.

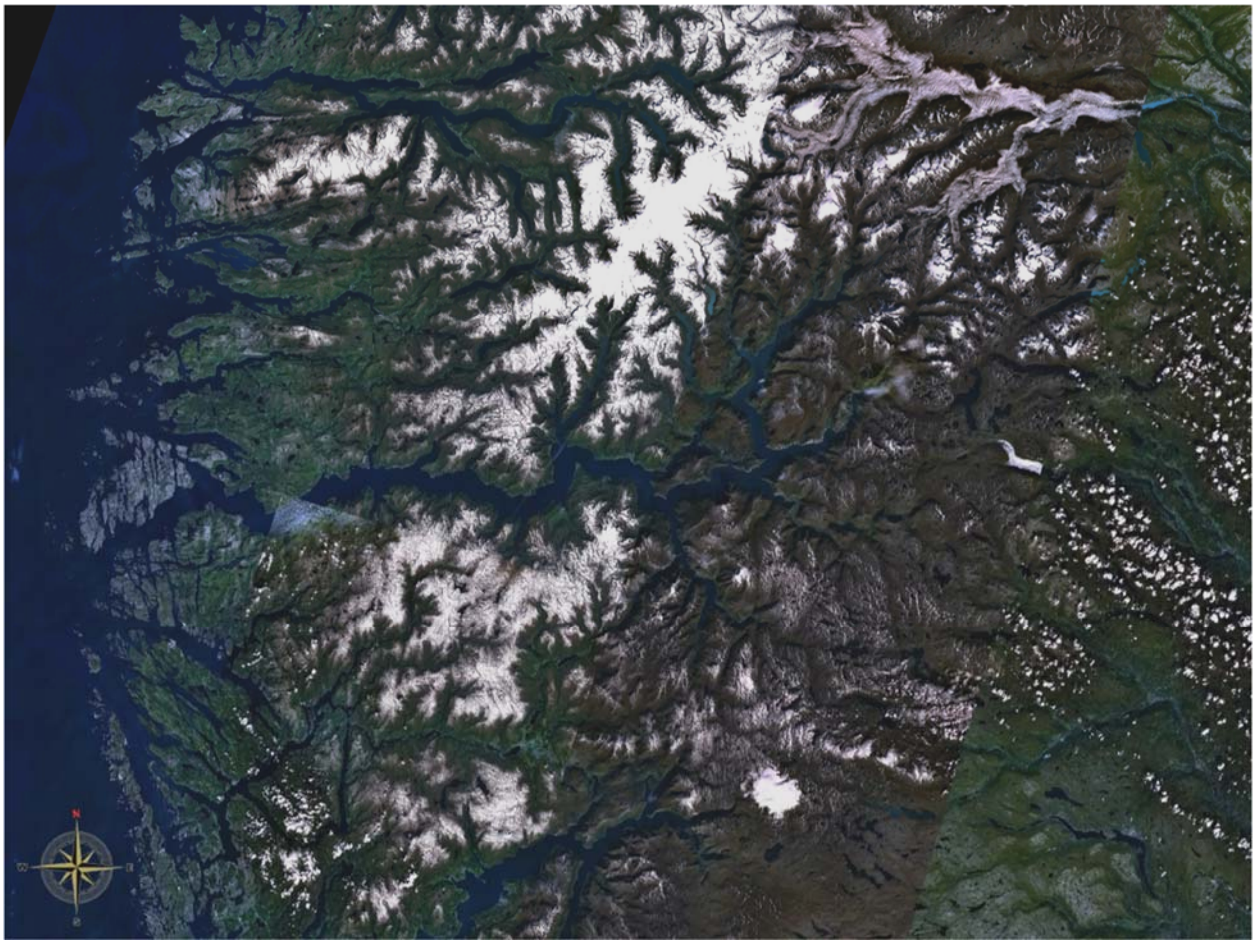

Figure 1.4. Satellite view of the Fjords of Norway taken by NASA Landsat imagery via Wikimedia Commons. These waterways approach a fractal pattern.

Fractional dimension is the simple, yet somewhat un-intuitive, concept that geometric objects like shapes and lines do not have to be described by wholesale dimensions. We are taught in school that lines are one dimensional, shapes like squares and triangles are two dimensional, and cubes and cones are three-dimensional. Imagine drawing a perfectly straight line. Clearly a one-dimensional object, right? But what happens when you take that perfectly straight line and draw a kink in it? Does that little kink suddenly mean that we are required to use an *entire* new dimension to describe the new space it occupies? In reality, it is entirely appropriate to describe the line as occupying something like 1.2 dimensions. It is also possible that your line could wind and wrap in such a way that it *does* need the lion's share of a new dimension to describe the space it occupies.

As far as names go, "Geometric objects with fractional dimensionality" is not very catchy so these objects are now popularly known as fractals: objects that look the same viewed either up close or at a distance. While these types of images have shown up on the dorm room walls and computer desktops of math nerds for decades, fractal structure can been seen every day in natural objects such as broccoli, the human vascular system, and possibly more relevant for this dissertation, the structure of the human brain. It is also thought that the physical connections that link neurons within the brain may also be organized in a in fractal-like way where statistical characteristics of the whole brain mirrors the characteristics of individuals parts at different scales; however, a precise mapping of every connection within the human brain is currently science fiction due to the combinatoric explosion of connections.

Using organisms with tiny neural networks such as the 302-neuron C. Elegans (White, Southgate, Thomson, & Brenner, 1986), we can start attempt to completely map out neural structure. With non-invasive imaging techniques, we can look at the relationships between structural and electrical activity to map interconnection of human brains. This approach of whole-brain mapping is currently the focus of the "Human Connectome Project" which has been touted by Vice President Joe Biden as an approach that "will lead to major advances in our understanding of how our brain circuitry changes as we age and how it differs in people with neurological or psychiatric illnesses" (Elam, 2011; Van Essen et al., 2013).

While the topic of the brain's structure is a new one, many studies have found that the brain has a 'small-world' structure. This type of network was first brought into the public by the oft-cited Watts and Strogatz paper from "Collective dynamics of 'small-world' networks". This paper introduced the concept that small world network structure has interesting properties that make it well-suited for certain type of optimization problems (1998). Brains, like any network, need to be connected in order to function but are limited because each connection has a cost. These costs may be due to the metabolic upkeep of neurons and supporting cells, information processing issues, and even the physical space these connections take up (Cahalane, 2013). Small-world networks are considered *efficient* because they balance keeping networks connected against the cost of connections. This optimization leads small-world networks to look like they have a fractal structure, and some scientists argue whether they *are* fractal or are simply *mostly* fractal (Gallos, Sigman, & Makse, 2012). This means that our brains themselves at the very least have a resemblance to the fractal structures seen all throughout nature. What effect does this have on our thoughts and behaviors? Can we find evidence of this structure in the things we do and create?

That idea is the focus of this dissertation. A lot is owed to the work of David Gilden, an astrophysicist-turned-psychophysicist who thought to model temporal structure of human reaction times. Without it, James Cutting and I would have been stuck retreading many of the basic statistics of film and I would not have been inspired to see how $1/f$ patterns influence our attention. What follows is an attempt to outline our early work on film, introduce a variety of $1/f$ modeling techniques, and finally introduce a new series of experiments which investigate whether 1/*f* structure really makes a difference to our attention and memory.

CHAPTER 2

Attention, Film, and 1/*f* Patterns

What is attention?

It is quite clear that human attention is a limited resource. While listening to a lecture or driving in a snowstorm, attention is something that must be paid. Other stimuli capture attention exogenously, pulling our brains towards stimuli in the world that might need our instantaneous reaction. Despite being a fundamental day-to-day experience, the description of what captures attention is more the domain of the artist, storyteller, and advertiser than cognitive psychologist. The slippery nature of attention was not unnoticed by pioneer psychologists. E. B. Titchener who noted that the study of attention was like a hornets' nest that, once approached experimentally, brings out a "whole swarm of insistent problems" (1908, p. 33). Titchener used the difficulties of defining attention as evidence that introspection is a valid approach, if not the most valid approach to the study of attention. To further bolster his argument, Titchener quoted the famous experimentalist Hermann Ebbinghaus as saying, "Die Aufmerksamkeit ist eine rechte Verlegenheit der Psychologie", translated by Titchener himself as "Attention is a real embarrassment for Psychology" (Ebbinghaus, 1911, p. 33; Titchener, 1908, p. 192). Other psychologists felt that the difficulties in studying attention could actually prove a common enemy that could ally the warring factions of introspectionism and objectivism (Kantor, 1922). However, the option to redefine or ignore the problem of attention proved to be a more attractive option for William James.

James neatly avoided having to define attention in detail, describing that that "Everyone knows what attention is" (1901, pp. 404–405). In the intervening 124 years since James' words were put into print, much has been done in order to describe the little we *do* understand about attention. Primarily based on studies in vision and audition, theories have shown that attention can act as a filter, allowing us to focus on specific stimuli in our environment either effortfully or automatically, as shown in the Cocktail Party Effect (Cherry, 1953). Psychologists have also debated whether the filter is something that is built into early stages of processing (Broadbent, 1958), attenuated before consciousness (Treisman, 1964), or functionally limited by memory capacity (Deutsch & Deutsch, 1963; Kahneman, 1973). Others have argued that 'filtering' may be an incorrect analogy because the brain can only process the stimuli it has already learned how to interpret perceptually (Hochberg, 1970; U. Neisser, 1967). These classic studies tend to focus on describing attention over the matter of a few seconds or shorter, typically introducing the limitations of memory through overwhelming the brain with input, reflecting William James' pronouncement that there is "no such thing as voluntary attention sustained for more than a few seconds at a time" (1901, p. 421).

Our knowledge of how attention can be sustained over time traces back to work done in World War II to test the perceptual performance of continuous radar monitoring. The Mackworth Clock is a device designed to test operator's vigilance: the operator watched the clock tick forward until the clock skipped an individual tick and the subject would press a button (Mackworth, 1948). This test would go on for multiple hours and each operator's performance was tested over time, typically falling off dramatically after thirty minutes. The Mackworth Clock is still a benchmark in the current field of psychology, being used in studies as diverse as determining the benefits of siestas (Sauter et al., 2013) to the emotional and cognitive deficits of dehydrated women (Pross et al., 2013). Mackworth fundamentally disagreed with James' take on attention and described the first fifty years of research on the subject of attention as "depressing reading – not only is it scanty, but it is also rather contradictory" (1948, p. 6).

Sustained attention has progressed in applied subfields interested in problems relating to human factors, education, and attention-deficit disorders while receiving less emphasis in the cognitive and perceptual subfields. The most mainstream impact of this line of work is the regulation of mandatory rest periods that are believed to increase sustained attention through sleep and/or boredom. Recent changes enacted as a result of the Colgan Air crash in Buffalo dictate that airline pilots can fly shifts of no more than nine hours before being grounded, and must have at least eight hours of uninterrupted sleep time before returning to work (Federal Aviation Administration, 2012; US National Transportation Safety Board, 2010).

Research has also been done to see whether taking breaks increases performance on tasks requiring sustained attention (Ariga & Lleras, 2011); however, these findings are contradicted by work that shows that simple disengagement is not enough to restore vigilance (Helton & Russell, 2012). A review of recent work has shown that meditation can increase attentional performance on certain tasks

(Sedlmeier et al., 2012).Typically, these studies involve quick tasks where attention does not need to be sustained, though a study by Valentine and Sweet showed that the ability to sustain attention is enhanced in those who meditate (1999). It is unclear whether those with enhanced attentional control are attracted to meditation or if meditation can improve an individual's sustained attention, but other studies have shown that training novices in mindfulness meditation appears to increase a variety of working-memory and attention tasks (Mrazek, Franklin, Phillips, Baird, & Schooler, 2013).

It may be obvious that sleep also plays a major role in sustained attention, but the function and process of the sleeping brain are still areas of active study. The concept of REM sleep was not conceived until the 1950's (Aserinsky & Kleitman, 1953), and it was posited as late as 1992 that REM sleep may only exist to keep the brain warm at night (Wehr, 1992). Sleep typically follows patterns beyond simple circadian rhythms that regulate day/night cycles, and oscillate with different stages of sleep, which come and go throughout the night. The foundation of this work was pioneered by Nathaniel Kleitman, who contributed to early work with a graduate assistant by attempting to adjust to a 28 hour day when deprived of sunlight inside Mammoth Cave (1939; Lavie, 1980a; United Press, 1938). Later, Kleitman published his research on the sleep habits and preferences of submariners aboard the USS Dogfish (1954). These studies, along with the advent of EEG as a practical research technique, lead to the development of a roughly 90-minute 'basic rest activity cycle' that regulates different stages of sleep (Kleitman, 1963).

Sleep scientists began branching out and investigating the possibility that the Kleitman's sleep rhythms may also be present when awake (Lavie, 1980b). These shorter oscillations are called ultradian rhythms, and have been posited as the driving force in a variety of arousal measurements including epinephrine level (Bossom, Natelson, Levin, & Stokes, 1983), reaction time (Conte, Ferlazzo, & Renzi, 1995; Gopher & Lavie, 1980; Lavie, 1980b), and pupil diameter (Lavie, 1979). As researchers continued to look, they found a number of other oscillations, including those longer than the Kleitman's 90-minute cycle, and named them 'slow ultradian rhythms' which may last for several hours (Lavie & Zomer, 1984; Manseau & Broughton, 1984). As these cycles did not have an obvious environmental correlate like circadian rhythms, the mechanism and function of these rhythms were brought into question. Some considered them a left-over function from sleep (Lavie, 1992) or possibly thermodynamic regulation (Lloyd & Stupfel, 1991). One particular study monitored subject performance at analytical and spatial tasks through eight hours of testing in an attempt to show that ultradian rhythms can be seen as a back-and-forth dominance of the brain's hemispheres over time (R. Klein & Armitage, 1979).

Intuitively, the fact that our perceptual ability waxes and wanes is often rolled into the area of physical performance, and in sports is referred to being 'streaky' or having a 'hot hand'. The study most associated with this phenomena was conducted by Gilovich, Vallone, and Tversky and examined the shooting performance of professional and collegiate basketball players using both video and experimental methods including a calculating serial correlation and running a Wald-Wolfowitz runs test (1985). Of the nine major players from the Philadephia 76ers 1980-1981 season, only one (Darryl "Chocolate Thunder" Dawkins), showed any evidence of previous shots affecting his percentage (Gilovich et al., 1985). It should be noted that Dawkins' performance during the season examined *was* atypical as he made 60.69% of his field goals, a performance that currently ranks 25th of all time (Basketball Reference, 2014). While the authors were originally reticent to expand this finding even to other sports, this study convinced many that the feeling that attention ebbs and flows is simply our brain poorly estimating probabilities in the real world.

As more instances of brain rhythms with varied oscillation rates appeared in the literature, it became more apparent that it is rare to find activity in the brain that *does not* follow some kind of non-random pattern. The emergence of complexity theory shed light on the nature of oscillations, pointing out that many systems in nature that appear stable are actually far from stable, fluctuating between alternative states that seem stable but can be provoked into chaos with seemingly small perturbations (Coullet & Tresser, 1978; Holling, 1973; Mackey & Glass, 1977). As the advanced mathematics of other fields were discovered and later adopted by psychologists, the ability to model these changes as more than simple oscillations can give us perspective into how the brain's natural rhythms influence our behavior.

1/*f* *patterns in attention*

Some of the first work recognizing that the fluctuation in reaction times was completed at Indiana University by showing subjects a luminance patch that was at visual threshold one of two ways: either at five-second intervals or in a self-paced manner (Verplanck, Collier, & Cotton, 1952). After analysis, it appeared that each subject's performance was not tied to their performance on the previous day, but rather how they had performed on trials that immediately preceded each test, as tested by looking at the correlations between trials taking place roughly a minute apart at Lags 1, 3, 5, 7, 9, and 20. This meant that rather than each

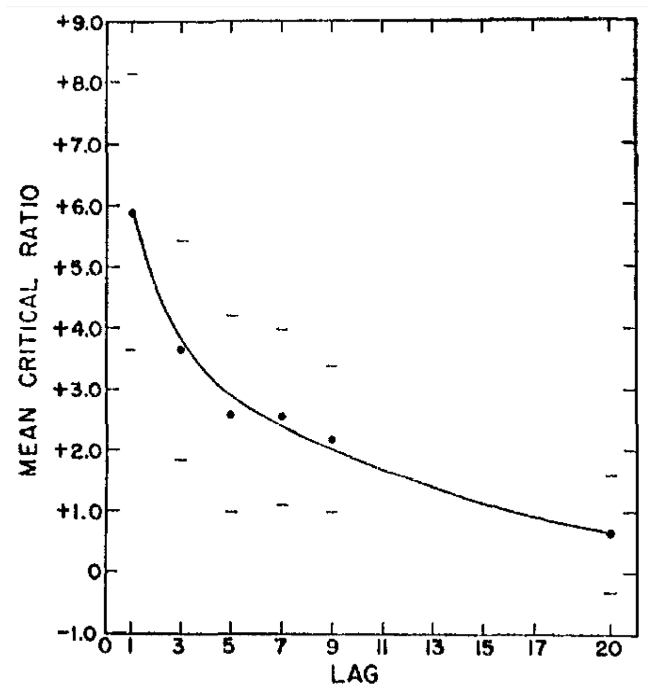

Figure 2.1. As lag increases, the probably of repeating an answer (e.g. being 'on a run') decreases. (Verplanck et al., 1952, p. 208)

individual reaction to the stimulus having *independent* variation, the changes in time were "dependent upon previous responses, or perhaps both are dependent upon a third variable which varies in time" (Verplanck et al., 1952, p. 281). Critiques of this article pointed out that the training procedure may have caused subjects to display a serial correlation, so a follow-up was published a year later showing the "lack of randomness" in reaction times (Verplanck, Cotton, & Collier, 1953, p. 13).

A spiritual successor to the Verplanck, Collier, & Cotton experiments was published in Science by Gilden, Thornton, and Mallon, who asked subjects to replicate either an interval in space or time by mimicking the observed delay in time or space (1995). This study broke ground by introducing the idea that it may be possible to model the fluctuations in reaction time (and possibly the mechanisms driving it) by using new timeseries analysis techniques adopted from the hard sciences and in Gilden's case, astrophysics. Instead of using serial correlation, which only tests the relationship between specific lags, Gilden used frequency analysis to look at the power spectra of the reaction times. Analyzing the power spectra is similar to looking at correlations across lags; rather than selecting a few lags that may seem important, comparing frequency components effectively looks at the influence of *all* lags within the signal. By looking at the relative strength of different frequencies across multiple experiments, Gilden concludes that reaction time can be modeled using two different types of noise: white noise and 1/*f* noise (1995, p. 1838).

White noise is a truly random change in a signal that has no correlation over time where the next value in the signal is uncorrelated with the previous value. This type of time series will not have any sort of regular frequency oscillation. 1/*f* noise is when previous values not only effect the probability of the next value, but also do so in a specific way. Simply put, values close in proximity are very influential, medium proximity are somewhat influential, and far-ranging values are slightly influential; this is much like the falloff found by Verplanck et al. (1952). The difference in a 1/*f* pattern is that the range and the influence are aligned in such a way that the relationship between them is maintained regardless of whether you're looking at the whole range or simply a segment of it.

By using power spectra across multiple experiments, Gilden was able to detect that while the reaction time errors of his subjects did have an element of pure randomness to them, they also had frequency levels that match a 1/*f* pattern (as seen in

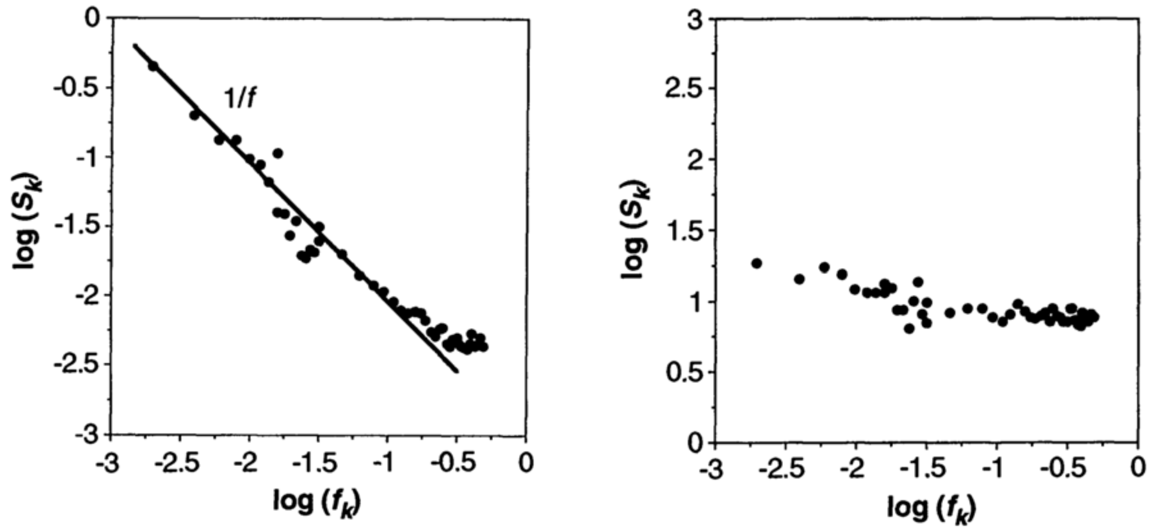

Figure 2.2: On the left, David Gilden's data show a 1/*f* relationship (straight line on a log-log plot) for a cognitive based task. Figure 2.2b shows a truly random relationship (white noise) that is emitted when subjects performed a simple reaction time without a 'cognitive' component. (Gilden, 2001)

Figure 2.2a). Interestingly, in an additional experiment, subjects were simply asked to press a button when a light turned on and did not show the same 1/*f* pattern in their reaction times (as seen in Figure 2.2b). With evidence for the 1/*f* pattern in the 'cognitive' task and not the simple task, Gilden concludes that "1/*f* noises arise from cognitive mechanisms that mediate the judgment of magnitude, independent of whether the magnitude exists in time or in space." (1995, p. 1838). This finding was expanded in a series of later experiments which display 1/*f* variation in mental rotation, a word/non-word judgment, serial and parallel searches, and priming, adding evidence to the claim that the emission of 1/*f* in reaction times isn't necessarily tied to time and space judgments, but appears in more general and varied tasks (Gilden, 1997, 2001).

Given that 1/*f* noise appears to be pervasive in most 'cognitive' tasks requiring attention, it stands to reason that we may be able to find this type of pattern in other places. Studies have found this type of structure in the amplitude of speech and music (Voss & Clarke, 1975) as well as the rhythm of Bach and Joplin (Levitin, Chordia, & Menon, 2012). Looking for this type of structure in film, which has only emerged as an art form within the last 100 years, could help shed light on what $1/f$ noise means for both attention and art.

Why study film?

Film is a great way to hold someone's attention. From rowdy children to airline passengers, it is hard to find a better way to get someone to sit still and stare forward for an hour and a half. Filmmakers work to capture attention with plot, character, location and special effects, but they also work with a set of rules that dictate how a film can be made. While it is unclear how the rules of music emerged, the rules of film composition were discovered through trial and error over the past century. Early filmmakers' intuitions about what would 'work' for audiences have been put on film, studied, and eventually adopted into the lexicon of film. Although many aspects of film appear to be developed differently based upon culture and style, some aspects of film remain remarkably stable. Rather than being culturally based, these aspects may tie in to more fundamental perceptual processes, which both limit and structure how a film can be put together.

At the dawn of cinema, filmmakers had a variety of technical limitations such as immobile cameras, dangerously combustible film, and short shooting time. In addition, filmmakers simply did not know what types of editing techniques their audience would be able to handle. It was feared that if viewers were exposed to a sudden change in camera position they would find it disorienting and aversive

(Bottomore, 1990). In order to circumvent this perceived issue Georges Méliès strung together shots using dissolves to ease the viewer from location to location (1899), much like a curtain between acts during a stage play. Other filmmakers believed a folk-perceptual idea that if a viewers blinked quickly during a cut, it might not be noticed and would lead to a smoother experience (Bottomore, 1990).

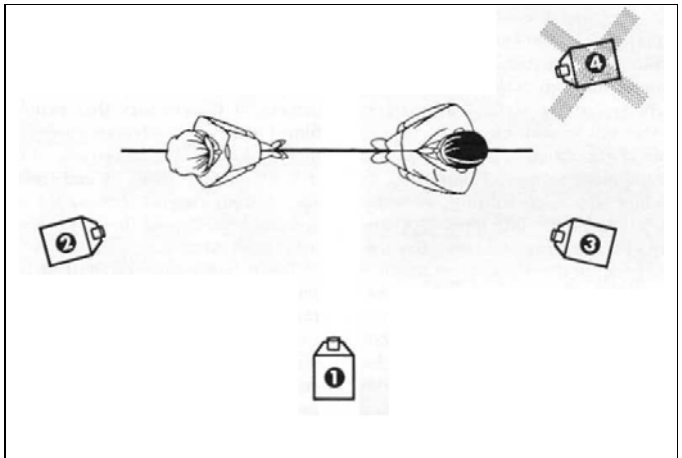

Figure 2.3: A diagram showing the basic structure of the 180º rule. While it is possible to cut between cameras 1, 2, and 3, cutting to camera four without reestablishing space will often confuse and disorient viewers (Mascelli, 1998, p. 110).

Filmmakers eventually discovered that the physical discontinuity introduced during a 'straight cut' is not aversive, but can often go completely unnoticed if placed correctly, even to observers specifically looking for transitions (Smith & Henderson, 2008). Not all cuts are equally transparent: relative location of people to one another and key objects need to be maintained. By keeping these relationships intact, filmmakers can feel free to move the camera around in order to get a clearer shot or simply to maintain visual interest. Later, this would become known as the 180° rule, which dictates that the transition from shot to shot must take place along one side of the 'line of action', the axis in which the action of the scene takes place (as seen in Figure 2.3). Other 'rules' have been observed and utilized informally by filmmakers for decades in the pursuit of a smooth and enjoyable perceptual experience.

It isn't clear why the system of editing developed in the early 20th century and continually refined works, however one explanation is that the shots in a film are put together in such a way that the brain can utilize standard perceptual processes to process the artificial stimulus of film (J D Anderson, 1998; Cutting, 2005; Gibson, 1979). Anthropological studies have shown that complete novices tend to re-discover the conventions of modern film even without training. Native Americans who had never seen a film were given the opportunity to shoot and edit their own movie without instruction other than how to use a camera and film splicer (Worth & Adair, 1970). Surprisingly, the Navajo settled on a film convention that respects spatial relationships the same way as mainstream film. These editing conventions do not appear to require any type of training. Pokot tribesmen who had not seen a film were shown two movies portraying the same events, one with editing effects (such as point-of-view shots) and another with no editing. Subjects were equally adept at describing what happened regardless of whether their version had the editing techniques (Hobbs, Frost, Davis, & Stauffer, 1988).

If the basic language of film can be understood without being explicitly learned, what in the natural world might prepare our visual system to understand it? Studies focused on people with different levels of film exposure from southern Turkey have shown that cues such as following gaze and assuming viewpoint still work despite the perceptual discontinuity of a cut (Schwan & Ildirar, 2010). Beyond simple physical layout, subjects are able to understand when transitions correspond with pictorial and causal relations, but struggle with understanding when transitions are conceptually driven (Ildirar & Schwan, 2014). Currently unpublished research shows that most issues associated with novice viewers can be alleviated by introducing diegetic sounds which further guide the viewer across perceptual discontinuities and allow even novice viewers the ability to predict the next shot in a sequence (Levin, Ildirar, Schwan, & Smith, 2014).

Cinematic cuts may be easily interpreted by the visual system, but that does not mean that the brain is passive throughout the transition. The brain knits together a continuous perceptual experience despite a physically discontinuous stimulus. This skill may have its roots in the same perceptual process that is used when the brain knits together a stable visual perception during and after the eye saccades, a concept that was touched upon slightly differently by the writer and director John Huston, who said in an interview that “All the things we have laboriously learnt to do with film, were already part of the physiological and psychological experience of man before film was invented…. Let me make an experiment – maybe you will understand better what I mean. Move your eyes, quickly, from an object on one side of the room to an object on the other side. In a film, you would use a cut. … In the same way, almost all devices of film have a physiological counter-part” (Bachmann & Huston, 1965, p. 10). Huston was correct about the suppressed visual input, but thought that the primary mechanism was a blink. Later research would show that while executing an eye movement the eyes are effectively blind for roughly 90 milliseconds (Volkmann, 1986) and our brains work to backfill the perceptual experience before we are consciously aware of the change (Thilo & Walsh, 2002).

Extensive eyetracking work by Tim Smith has shown that the cinematic system of continuity adeptly manipulates and guides viewers’ gaze patterns around film through three fundamental stages: attending to a shot, cuing attention across a cut, and matching expectations after a cut (2012, p. 17). These steps do not require conscious effort, but with every shot, our visual system is asked to interpret a number of questions within each image including which characters are present, where they are, and what is going to happen next.

Although not formally expressed, filmmakers have been well ahead of the curve in understanding how to manipulate our attention and keep us glued to the screen. Filmmakers been working to capture and keep our attention long term for the last century and have passed on knowledge, both explicitly in instruction and implicitly in their work, about how to capture, keep, and manipulate human attention over longer stretches of time. It stands to reason that we may look to film to see if we can tap into some of the collective insights produced by filmmakers in their pursuit.

Film Selection

In order to determine whether filmmakers have changed their techniques over time we need to select a large enough number of films that are at least roughly representative. Rather than attempt to study every type of film produced in the world, we initially focus on a sample of 150 popular films shot in the 'Hollywood-style' from 1935 to 2005 (Cutting, DeLong, & Nothelfer, 2010), with an additional ten films added in a later paper (Cutting, Brunick, DeLong, Iricinschi, & Candan, 2011). 1935 was chosen as a starting point because films with sound had become the norm and many were available in digital form. We expect that other studio systems and film movements across the world (such as French New Wave, Dogme 95, and Bollywood film) will not perfectly adhere to any general findings from our Hollywood database, and are ripe for examination from the film community at large. Our choice to study Hollywood Film stems from the fact that studios have financed and produced films in a somewhat similar manner for a long swath of time, the films have broad commercial appeal, and are easy to obtain digitally.

Individual films within each year were selected using ratings and genre classifications from the IMDb website as well as box office data when available (IMDb Inc., 2014). For each year, we attempted to select films that fit each of five major genres: action, adventure, comedy, drama, and animation, but trends in film sometimes manifest in such a way that the most successful films of the year may have excluded one or more of these genres. Each film was ripped from NTSC DVD sources and transformed into a 256x256 pixel movie with a constant rate of 24 frames per second.

Shot Detection

One of the most studied and commented-upon changes in film has been the pace of editing in film, typically by calculating average shot length. By dividing the length of the film (typically excluding titles and credits without any story-related content) by the number of shots within the film, one can get a general idea of how often a transition occurs. Film theorists simply have to count every time a cut appears onscreen, but as stated before, we have a tendency to miss cuts even when we are explicitly looking for them (Smith & Henderson, 2008). It is also possible to go through the movie frame-by-frame, however this method can be prohibitively time consuming for the individual researcher.

A generally accurate and time-efficient way to detect cuts is by using an algorithm to automatically detect when inter-frame changes pass a certain threshold. While cuts between static images are typically quite different statistically and easy to detect, other transitions such as fades and dissolves may only change gradually over the course of several dozen frames, making detection difficult. At the moment, there are limitations to the performance of such algorithms. In 2007, the TRECVID competition featured a shot boundary detection task between several of the world's leading academic and professional teams like IBM Research, the Imperial College London, the Motorola Media Research laboratory, and the Indian Institute of Technology (Smeaton, Over, & Doherty, 2010). Overall, even the best algorithms were only able to recognize straight cuts 93% of the time with a 5% false alarm rate. Detection is worse for gradual transitions like dissolves and fades, where only 78% are detected with an 11% false alarm rate.

In order to detect shot boundaries in our film dataset, we created an algorithm in MATLAB (The Mathworks Inc., 2007) which divided the 256x256 frame into an 8x8 grid. Luminance histograms from the three color channels (when available) within each section of the grid were compared to the corresponding grid section histograms for both the previous frame as well as the subsequent frame. Using the third derivative of the change between frames (jerk), the change between each grid section, and covariance between the grid sections we were able to avoid false-alarms. In addition, by modulating our response by covariance, we could make sure that any single major change in a grid would be tempered by whether that change was occurring over the entire frame. Detecting gradual transitions was attempted by looking for a monotonic change in luminance across at least six frames (to detect fades) and by looking to see if the distribution of pixel-level values were Gaussian (to detect dissolves). Neither system for detecting gradual transmissions is recommended for future use.

After the algorithm detected a list of candidate cuts researchers used a GUI interface to check that each candidate was correctly labelled as a transition as well as flag where cuts may have been missed. By elevating the false-alarm rate and allowing researchers to quickly reject candidate cuts, the system was able to detect 99% of all transitions with a 0.2% false alarm rate when compared to manually coded, frame-by-frame analysis of “Revenge of the Sith” and “Spies Like Us”. Following every stage of the encoding process, each film would require, at minimum, twelve hours of computational time and four hours of human intervention. This translates to roughly 1800 hours of computational time combined with 600 hours of checking the data in order to accurately detect as many transitions as possible.

In studies conducted after 2009, the hybrid computational/human approach was streamlined in favor of having research assistants hand code all films, reducing the computational bottleneck. The data was once again revisited to check for any coding inaccuracies before publication of a paper featuring fades and dissolves, (Cutting, Brunick, & DeLong, 2011). This process involved looking for any lost transitions between those previously flagged. Once all transitions were isolated, they were labeled as a specific type (cut, fade out, fade in, dissolve, wipe, and other).

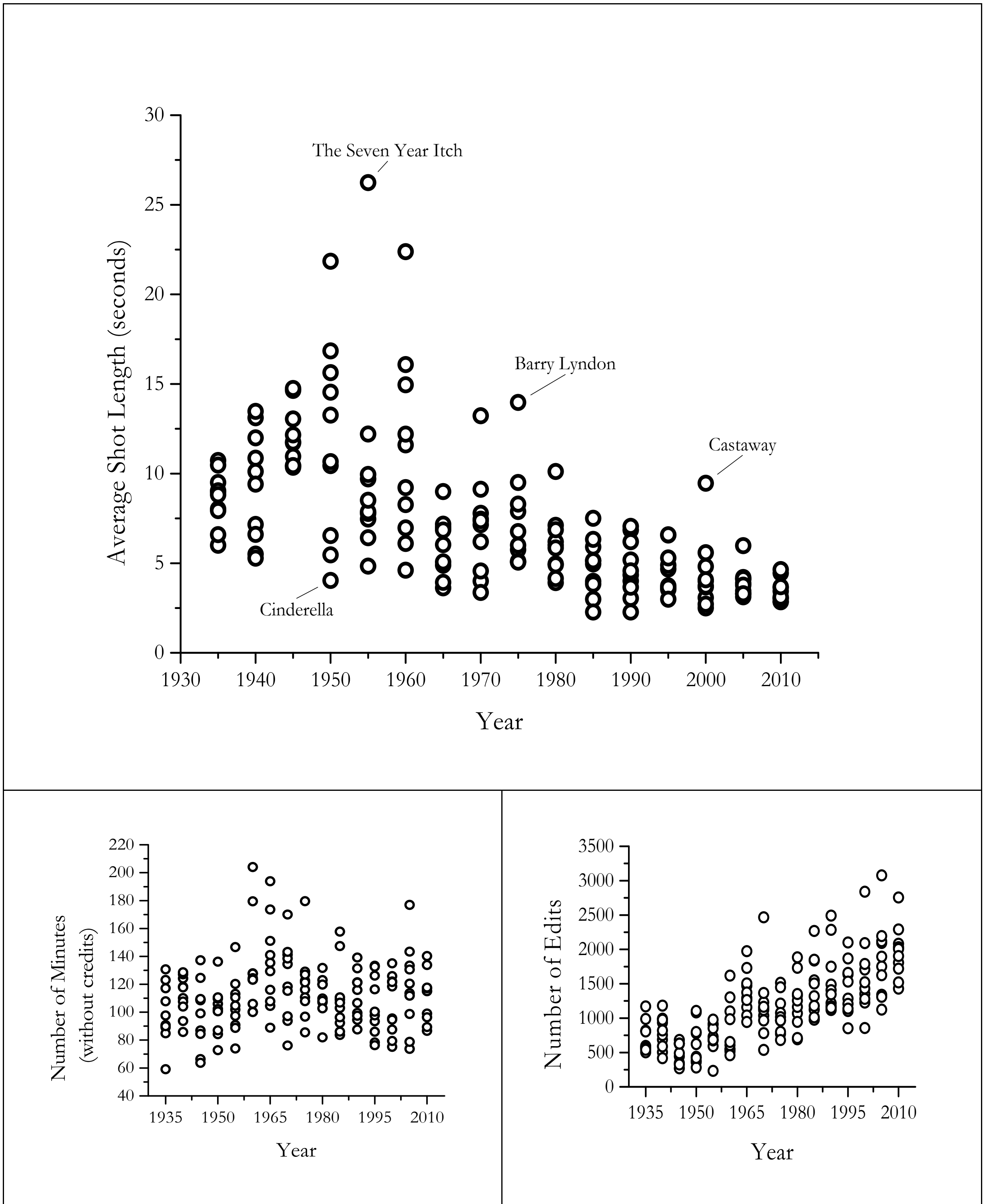

Figure 2.4: The upper panel (Figure 2.4a) shows the decrease in average shot length over time. The bottom two figure elaborate on why this is the case, showing that film length has stayed somewhat stable (Figure 2.4b) but the number of edits has increased (Figure 2.4c).

Accelerating Editing

Film scholars have noted that shots were getting shorter by simple observation (Bordwell, 2006) and through a large review of over 7,000 films (Salt, 1992, 2006). This type of study has also migrated online where film researchers can find both software and a database of publicly collected shot lengths totaling over 14,000 entries, although it is not known how accurate these submissions are (Tsivian & Civjans, 2006). The literature agrees that shot lengths have decreased dramatically since the 1930's. Our data replicates the findings, showing a decrease in overall shot length from 1960 to today but also showing a great deal of variety in the 1930's through 1950's (as seen in Figure 2.4a). This change is not due to movies becoming shorter over time (Figure 2.4b) but the number of edits increasing (Figure 2.4c).

The reason for the decrease in shot length is not immediately clear. David Bordwell has posited that decreased shot lengths, along with "the bipolar extremes of lens length, a reliance on tight singles, and the free-ranging camera" are part of a general style called *Intensified Continuity* that has increasingly influenced filmmaking since the 1960's (Bordwell, 2006, p. 137). Folk ideas of why shot lengths continually decrease typically try to explain the change using

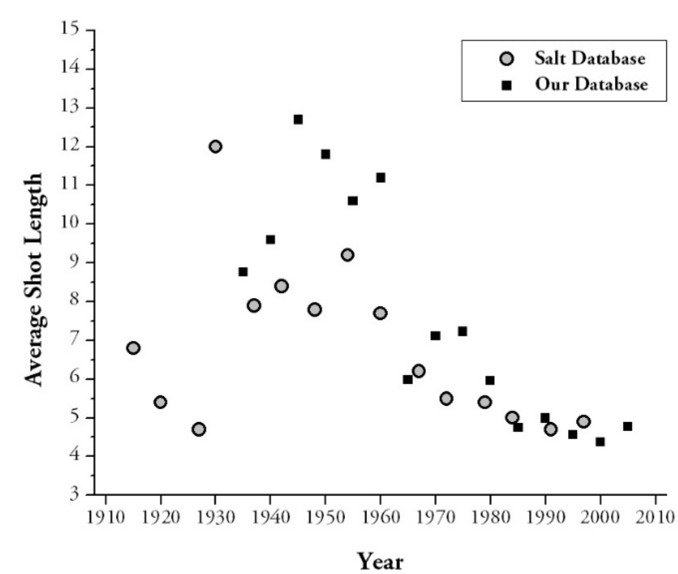

Figure 2.5: While our database ends at 1935, Barry Salt has looked at films from the silent and early film period. For the time period we have observed, our sample appears to line up well. (Salt, 2006)

technological innovations in editing and filmmaking. Surprisingly, the trends currently experienced in film are part of a more continuous change. If the adoption of editing technology were to blame, we would expect a spike as studios transitioned from film to digital editing. A more reasonable idea is that filmmakers now shoot more takes, allowing the editor more freedom to make edits in a variety of places.

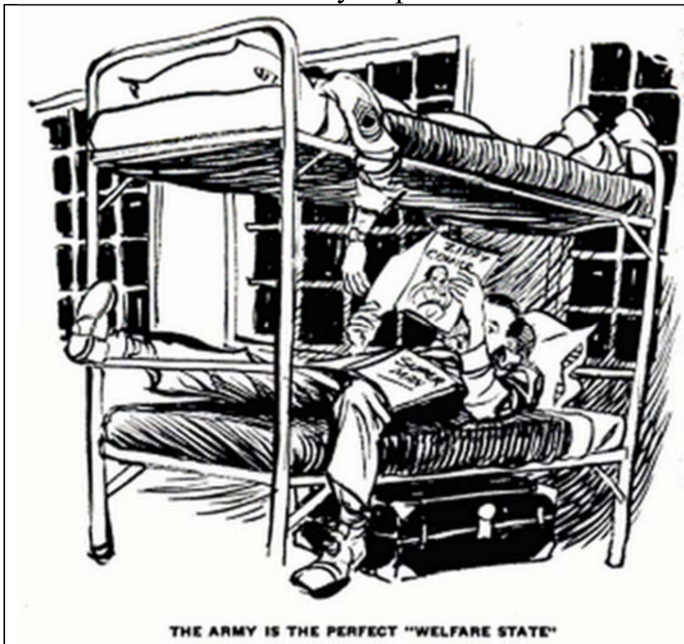

Figure 2.6: As *The Greatest Generation* returned from war, they were criticized by their parents as lazy and addicted to unhealthy media such as comic books. It appears younger generations cannot please their parents, even after defeating Hitler (Mauldin, 1950).

Popular media pundits often champion the argument that the current generation's minds have been so damaged by fast moving new media such as music videos, the internet, or video games that they can no longer pay attention to anything lasting longer than a few seconds. This movement is little more than an ephebiphobic tantrum, and has been featured prominently in American discourse for at least the last century. After winning World War II, the parents of the Greatest Generation were quick to voice their concerns that life in the army had made their young men lazy, as they preferred to stay at home and engage in reading the often immoral comic books, a criticism lampooned in Figure 2.6. Years later, Frank Sinatra would slam the movie and music industries, saying "My only deep sorrow is the unrelenting insistence of recording and motion picture companies on churning out the most brutal, ugly, degenerate, vicious form of expression it has been my displeasure to hear – I refer to rock 'n' roll. It fosters almost totally negative and destructive reactions in young people. It smells phony and false. It is sung, played and written, for the most part, by cretinous goons, and its almost imbecilic reiterations and sly, lewd – in fact, plain dirty – lyrics make it the martial music of every sideburned delinquent on the face of the earth." (Turner, 2004, p. 104).

Shot Distribution and the Lognormal Distribution

The precise identification of transitions in our database allows us to also investigate beyond the single parameter of average shot length and start to talk about other aspects of the shot length distribution. As most film scholars were simply counting the number of cuts in a film, it was not clear whether the mean was actually a good measure of central tendency. The first person to investigate the distribution of shot lengths in a film was Barry Salt who suggested films may fit a Poisson distribution (1974), but later amended his analysis to claim that the Lognormal distribution may be a better fit (1992).

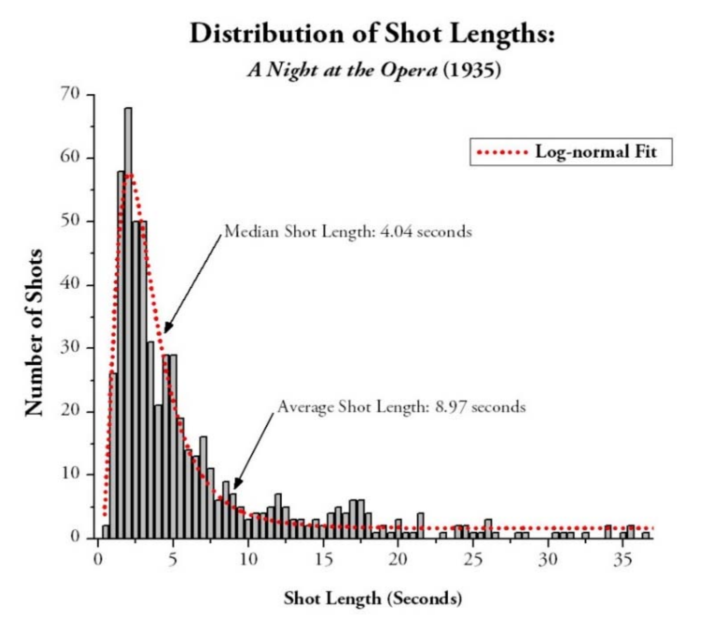

Figure 2.7: A histogram of each shot in the 1935 film, A Night At the Opera with a log-normal fit overlaid. The shots do not perfectly follow the lognormal distribution; however, the shape does appear to be a fairly close approximation.

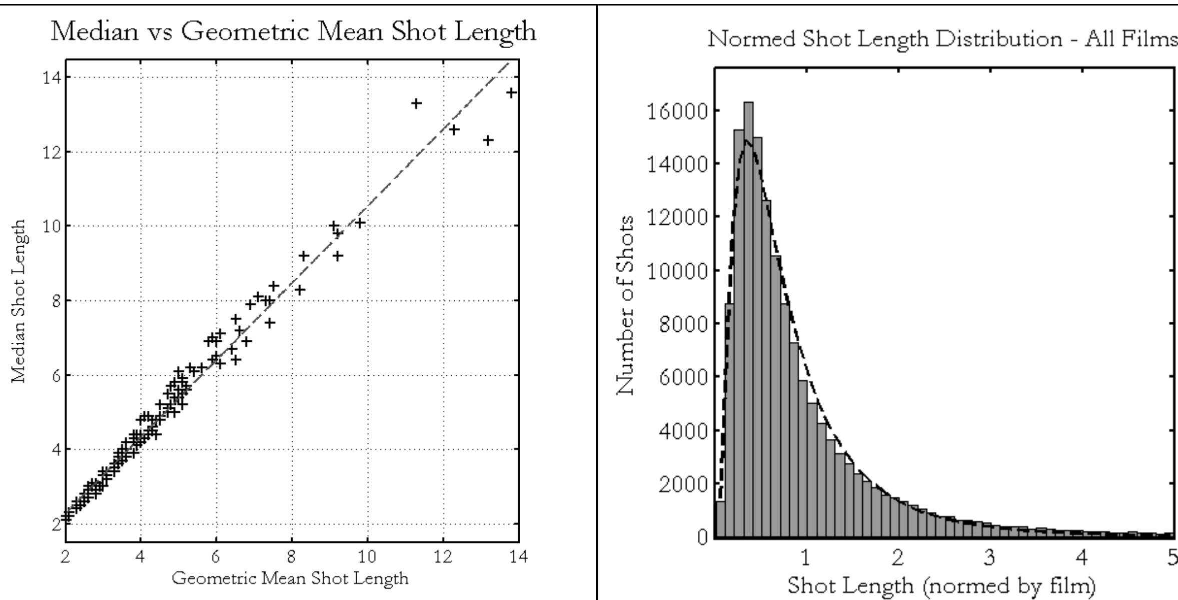

Figure 2.8: In the figure on the left, each film's mean and median shot length are compared to show a strong relationship between the two. In the figure on the right, each film was normalized to have the same average shot length, and then all shots in the database were put into a histogram to check whether a lognormal distribution is a good approximation of the global distribution

As pointed out in a previously published chapter, utilizing the arithmetic mean to characterize a highly skewed distribution can lead to inflation of the mean through some very long shots (DeLong, Brunick, & Cutting, 2012). The average shot length is actually much longer than most shots in the film, as seen in Figure 2.7.

This presents a problem: if the distribution of shots in a film are highly skewed, it means that our estimates of decreasing shot length may be driven by the elimination of extreme-long duration shots rather than a shift of the entire distribution. Using the median shot length, while more difficult to collect, should provide a more accurate measure of central tendency than the mean. While it is true that shot lengths form a highly skewed distribution, if they are *reliably* highly skewed to the same degree, the relationship between the mean and median should remain stable over time.

Figure 2.8a shows that the relationship between median and mean for our dataset is closely related; suggesting the overall skew of each film is similar.

A deeper examination of fitting the lognormal distribution to shot length data has shown that not *every* film in our database adheres perfectly to a lognormal distribution (the shot distribution from A Night at the Opera in Figure 2.6 is actually considered a poor fit), most are strikingly close (DeLong, 2013). When normalizing across the entire database, it becomes clear that the lognormal distribution is, at the very least, a convenient and appropriate approximation for shot lengths in our database, as seen in Figure 2.8b. Another surprising aspect of the shot length distribution's approximate lognormality is that it did not gradually evolve over many decades like many of the characteristics of film our database, but has been present and stable in film at least to the point where our sampling begins in 1935 (DeLong, 2013). It is not clear why films follow this pattern so closely.

The lognormal distribution has been mysterious in its ubiquity, not unlike the current status of the $1/f$ pattern. The lognormal distribution was originally identified by pioneer psychophysicist Gustav Fechner (1860) but was later recognized in art when an entomologist was reading a paper on the lengths of sentences in literature. He recognized that the graphs of sentence length had an uncanny resemblance to some of the data he had collected on insects. After analyzing 600 sentences from three books, he found that if sentence lengths were binned using a logarithmic scale (rather than linear scale) that the distribution appeared normal and symmetric (Williams, 1940). Aspects of the lognormal distribution have been described in-depth by economists (Aitchison & Brown, 1957) but what type of underlying mechanism may cause this type of pattern remains unclear, at least when compared to sources of variation that are thought to give rise to the Gaussian distribution (Limpert, Stahel, & Abbt, 2001).

Lognormal distributions appear in situations where there are lower boundaries that variables cannot pass. One such example is the time between contracting a disease like chicken pox or salmonella and experiencing symptoms. Diseases cannot be contracted before exposure, almost never present at once, typically last for a median amount of time (14 days for chicken pox, 2.4 days for salmonella), and will occasionally drag on for much longer than typically experienced (Sartwell, 1950). Taking into account the difference between distributions becomes very important when modeling epidemics, as a Gaussian distribution would critically underestimate the amount of patients exhibiting a very long latency period.

The lognormal distribution is also necessary for modeling reaction time experiments. Franciscus Donders, an ophthalmologist and friend of Hermann Von Helmholtz, was the first to collect data in the lab confirming the belief that task difficulty moderates the speed of response (1867). Reaction times entered into the mainstream within the work of Francis Galton, who collected data en-masse to determine the range of human genetic variability and capability both physically and psychologically. While Galton is known for pioneering comparative and population techniques in psychology, he also promoted the less palatable concept of eugenics, a term he coined in the same volume he introduced his reaction time measures (1883). Reaction times have been continually used since Galton’s work; however, it wasn’t until later that psychologists realized that reaction times are far from normally distributed and began, as an ad-hoc reaction to skewed distribution, regularly log-transforming reaction time data (Schlosberg & Heineman, 1950, p. 241). Decades later, researchers would attempt to understand why reaction times may be lognormal in the first place, settling on the fact that that distribution may be caused by three factors: (Ulrich & Miller, 1993, p. 523)

1. A normal distribution which is run through a multiplicative process
2. A variety of independent, random variables
3. Both

It is possible to theorize why these types of distribution may emerge from neural systems. Neurons are, by necessity and/or function, noisy agents in a complex system. The temporal variation from entire populations could be Gaussian in nature but before being observed as reaction times, are run through a process that is multiplicative in nature where multiple feedback loops are generated like decision making or motor planning. Alternatively, it is also possible that the variation generated by our brains is done at multiple scales simultaneously. While the variation of each processing stage is Gaussian in nature, the (somewhat) serial organization of information processing would mean that each level would contribute variation at different scales, possibly depending upon the size of the neural population. If this were the case, we should be able to understand more about the organization of a neural system, simply given the distribution of reaction time responses.

It is more difficult to theorize why shot lengths have followed a reliable lognormal distribution throughout the decades of our film database (DeLong, 2013), as they are generated by filmmakers who have a great deal of freedom to craft shots as they see fit. Pushing a button to respond during a reaction time task may require a cognitive effort, but filmmakers are engaged in the metacognitive exercise of trying to gauge how audiences will react to an edit. The fact that the lognormal distribution is present when trying to be ‘understood’, in film, literature, or language, may be a meaningful feature of our how our brains process information or is simply a pattern so ubiquitous that it may be hardly worth mentioning.

1/f patterns in Hollywood Film editing

When viewed on an unlabeled graph, it is difficult to tell the difference between a series of shot lengths and a series of reaction times. Because of the similarities between the datasets, we can use the exact same toolbox to process both. For our studies, this will primarily mean utilizing the techniques developed by David Gilden (1995, 1997, 2001). On the outset, reaction times are typically much shorter than shot length for films, especially when comparing older films with higher shot durations. This gap is shortened if we take into account that many experiments include an intertrial interval that may last for several seconds. Regardless, the difference in average length does not have an effect on our procedure.

The easiest feature to pick out of the graphs is the approximately lognormal distribution of both shots in film and reaction times– mostly short and medium entries, which are occasionally punctuated by an extremely long trial, and shot. In film, this often takes the form of a long-duration establishing shot that may feature a crane or helicopter zooming over a landscape. In this case, the filmmaker exogenously chooses the pause for you. While participating in reaction time tasks, this may take the form of ‘spacing out’ or mind wandering between trials. Something different is happening in a reaction time experiment, where an individual exhibits very different reactions to a same or similar stimulus due to endogenously driven factors. Most experimenters have an intuitive appreciation for this fact and, in order to combat fatigue, will put short breaks between trials (inter-trial interval), medium breaks between blocks, and long breaks between experiments. The typical experimental script in psychology was not dictated based on a series of rigorous methods, but was developed through trial and error during the pioneer phase of psychology, and has been passed down from professor to student.

Spotting whether a timeseries 'looks' $1/f$ is more difficult than spotting if it is lognormal and the two should not be confused. It is important to note that that while lognormality and the $1/f$ distribution often appear together in this dissertation, it is possible to have a $1/f$ timeseries that is *not* lognormal and vice versa. A $1/f$ pattern is not a distribution of shots, but a relationship between shots over many scales. Given that we cannot easily recognize these relationships visually in the raw timeseries data, it makes sense to transpose the data into the frequency domain, which will allow for more straightforward graphing and analysis.

Gilden's analysis is based upon finding the relationship between power and frequency, plotting the line that best fits that relationship on a log-log graph, and then reporting the slope of that line as α, which stands in for an indirect estimate of dimensionality. As mentioned in Chapter 1, a flat slope suggests no pattern between different frequencies (white noise), a slope of -1 is a $1/f$ pattern (pink noise), and a slope of -2 is a random walk pattern (Brown noise). The Gilden analysis also attempts to estimate (and correct for) any white noise that may have been mixed into the signal.

This can be seen in situations where the frequency stops decreasing and 'flattens out' due to noise.

Another parameter introduced in Gilden's analysis and relevant to our technique is how many points along the frequency/power axis should be included when fitting the slope. As seen in Figure 2.4c, modern films may have 2,000-3,000 shots while our shortest film, The Seven Year Itch, only has 233. Shorter films simply will not contain information at the longest timescales because the pattern is not present or the number of samples needed to detect a pattern will be too low. For each film, we attempted to balance between fitting as many points as possible and maximizing the quality of the fit. While this is not the only way (or perhaps, most optimal way) to estimate dimensionality, the accessibility of the model, prevalence in the literature, and guidance from David Gilden made his technique a sensible first choice for our data.

As can be seen in Figure 2.9, the outcome of our analysis is not as obvious and robust of an effect as shot length through the years; however, it is easy to interpret a couple major movements. Earlier films have a larger range of α, which is the parameter of dimensionality that is fit by the analysis. The fact that these films have shorter shot lengths may mean that our analysis is open to more random variation than with later shots but could just as easily mean that filmmaker's practices were not as formalized as later years and were open to more variation. It is clear that decades are not particularly restricted in their range of α, as many years include films that span the entire gap between white noise and pink noise, with some going past pink noise (where α=1) and moving closer to being modeled as Brownian noise.

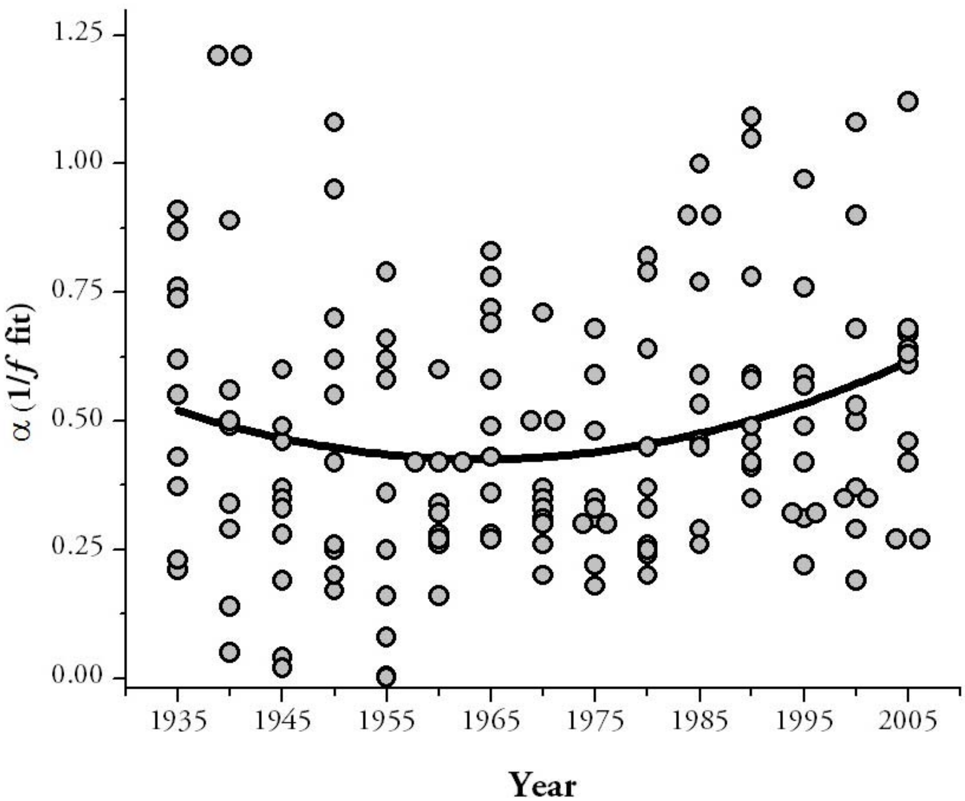

Figure 2.9: Over time (and especially since the mid 1960's) it appears that films are becoming more 1/*f* in their structure.

From 1935 to 1955, there is a great deal of variation in the data and it looks as if the films may actually exhibit less of a 1/*f* pattern over time. While it is possible that the analysis is locking into some sort of downtrend present in the sample, it may be that the relatively small sample of films with a small number of shots is influenced by random variation. It is also possible that our sample began capturing films on the downswing of a cyclical 1/*f* pattern that will unfold over the next century or so, but this seems even more far-fetched.

From 1960 to present day, we can see that films in our sample appear to fit a 1/*f* pattern over time more closely. The direction is modest yet reliable, and was

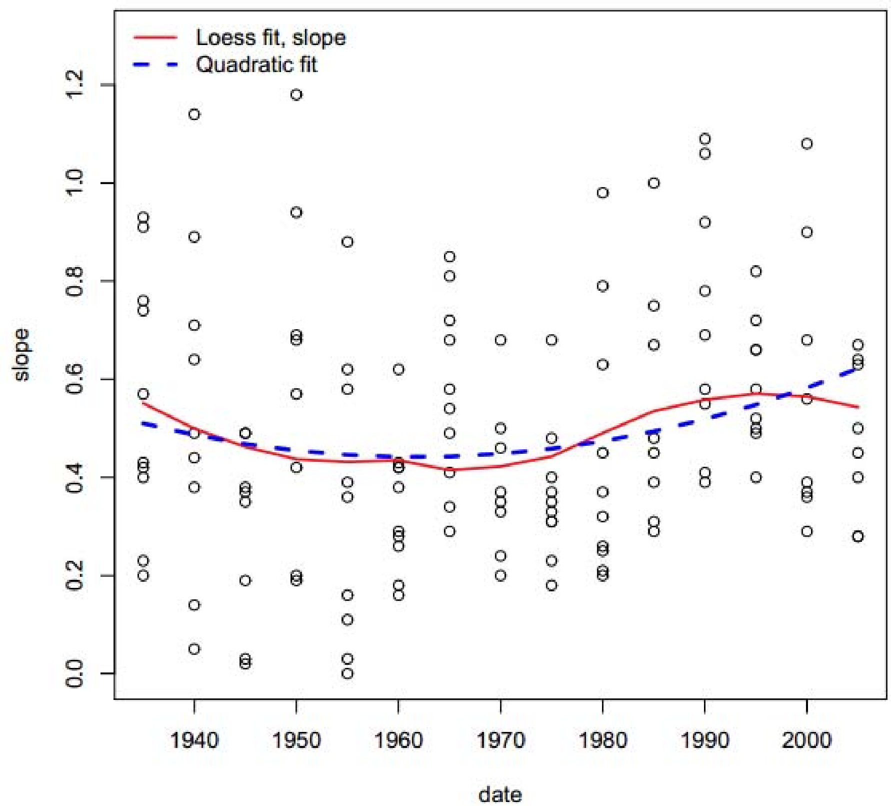

Figure 2.10: Using our data, which has been made publicly available on the Cinemetrics website, Mike Baxter re-fit the data with a LOESS fit. (Baxter, 2013)This fit shows a different pattern that the quadratic fit used our original paper, showing no real increase other than during the 1960's to 1990's.

reinforced after data from 2010 was added to the database (Cutting, Brunick, DeLong, et al., 2011). Film scholars will note that the late 1960's to early 1980's is considered an important era in Hollywood cinema, as smaller films with auteur direction (like "Easy Rider" and "Raging Bull") inspired the renaissance known as 'New Hollywood'. This has been characterized as a "strident stylization" where faster cutting gained mainstream adoption (Carroll, 1998). Whether these stylistic changes have a contribution to the mathematical differences we observe in our films is unknown, but may make contextual sense considering the stylistic changes involved in intensified continuity (Bordwell, 2006). This style grabs attention through removing almost all parts of a film sequence not needed for viewer understanding while retaining the correct rules governing traditional filmmaking.

The data reported in our original paper was fit using a quadratic fit which appears to describe the initial change from 1935-1955 as well as the increase from 1965-2005 (Cutting et al., 2010). It is also possible to do a more straightforward linear fit, which remains significantly positive but projects the increase in the direction of the $1/f$ progressing very slowly. Mike Baxter has proposed that there is realistically no overall trend in the data (except for possibly during the 'New Hollywood' years) and that a LOESS fit is a better representation, as seen in Figure 2.11 (Baxter, 2013). LOESS fits are good for our data because we did not propose any particular type of change before collecting the data and our linear and quadratic fits are not accounting for a lot of the changes in our noisy data. Unfortunately, LOESS fits are highly susceptible to outliers and our dataset is full of them. In a separate response, James Cutting proposed that LOESS might actually be overfitting the data. While LOESS fits are not directly comparable to linear and quadratic fits, a parametric approximation of the LOESS fit does not gain enough predictive power per parameter to justify using it.

Rather than debate the overall fit from one individual method, it is helpful to look at how other methods of detecting $1/f$ patterns in our data may shed light on the issue. While reporting results from the first 150 films, we included a separate analysis of the data based upon autoregressive techniques in hopes to supplement the Gilden analysis (Cutting et al., 2010). The following sections of this chapter will investigate several other techniques for analyzing data from our shot lengths database and comparing their accuracy and approach.

Wald-Wolfowitz Runs Test

The Wald-Wolfowitz runs test is a non-parametric technique developed by two Europeans who relocated to the US between World Wars I and II. Wald emigrated after facing religious persecution, eventually meeting Wolfowitz while in New York. During the war, Wald pioneered statistics which would help determine which areas of planes needed to be reinforced based upon which areas are typically hit, how many times a plane is typically hit, and which type of ammunition is most likely to be used by the enemy (Wald, 1980). With bitter irony, Wald died in a plane crash in 1950, leaving Wolfowitz to defend and promote their joint work. A year after Wald's death, Wolfowitz left NYU to became a professor of mathematics at Cornell University and would help found the field of information theory. His son, Paul Wolfowitz, grew up in Ithaca, graduated from Cornell, and would eventually become a core architect for the Bush Doctrine and invasion of Iraq (Bacevich, 2013).

The runs test works by choosing a value in the middle of a distribution (typically the median value) and then counting the number of consecutive samples below or above this value. If the sequence is truly random, the predicted number of runs should follow a normal probability distribution. If there are less runs than predicted, then it is safe to assume that there are clusters of similar values and the sequence is not random (Wald & Wolfowitz, 1943). If there are more runs than predicted, it also means that the sequence is nonrandom, but is alternating below and above the middle of the distribution in a set pattern.

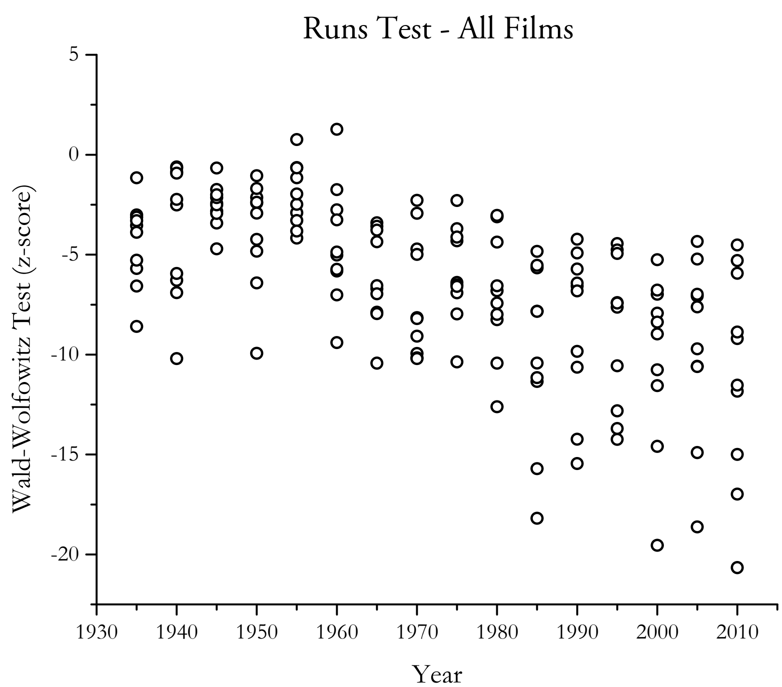

Figure 2.11: The results of the Wald-Wolfowitz runs test on each film in our database. Films appear to exhibit less runs than expected over time, however this result may be influenced by the raw number of samples in later films.

In our database, only sixteen films are not significantly different than chance, suggesting that most films in the database are constructed in a non-random manner. Rather than simply compare significance levels, it may be more instructive to look at the z-score output of the test, as seen in Figure 2.12. Earlier films appear to be closer to a random sequence than later films, but there is a consistent trend for all films to display *fewer* runs than predicted.

This method provides a good first step to see that sequences are not purely random; however, it cannot model the type of non-randomness in the sequence. The number of samples also may influence the output of the method, which will be larger in later films with more shots.

Autocorrelation and Partial Autocorrelation Function

The invention of autoregressive techniques is primarily credited to Udny Yule, a Cambridge statistician whose papers in the 1920's mark a milestone in the field of timeseries analysis (1921, 1926, 1927). Yule had previously demonstrated that two non-correlated harmonic timeseries could actually be seen as correlated if sampling wasn't done correctly (1926). However, Yule was still trying to develop methods to determine whether a system was simply changing over time randomly, or if it actually followed some sort of long-memory. In order to check this, Yule worked with odd datasets including wheat prices in Western Europe from 1545 to 1844, recordings of rainfall in at Greenwich from 1815-1924 (J. Klein, 1997), and the number of relative sunspots which have been accurately collected from 1848 until today (Wolf, 1848).

Yule's developments led to a method in which a timeseries can be serially correlated with itself in an attempt to expose, not only whether there are repeated patterns in a timeseries, but also which scale they occur. This method can be thought of as a simple correlation between a timeseries and itself, with the timeseries shifted by some number of samples (known as lag), and was previously outlined when talking about the performance of NBA players (Gilovich et al., 1985). Correlating the timeseries with itself without shifting (Lag 0) will always show a perfect relationship. Comparing the timeseries with a version of itself offset by a single value (Lag 1), could either show no predictive value (indicative of white noise) or show some amount of predictive power (indicative of some kind of noise that *is not* white). We

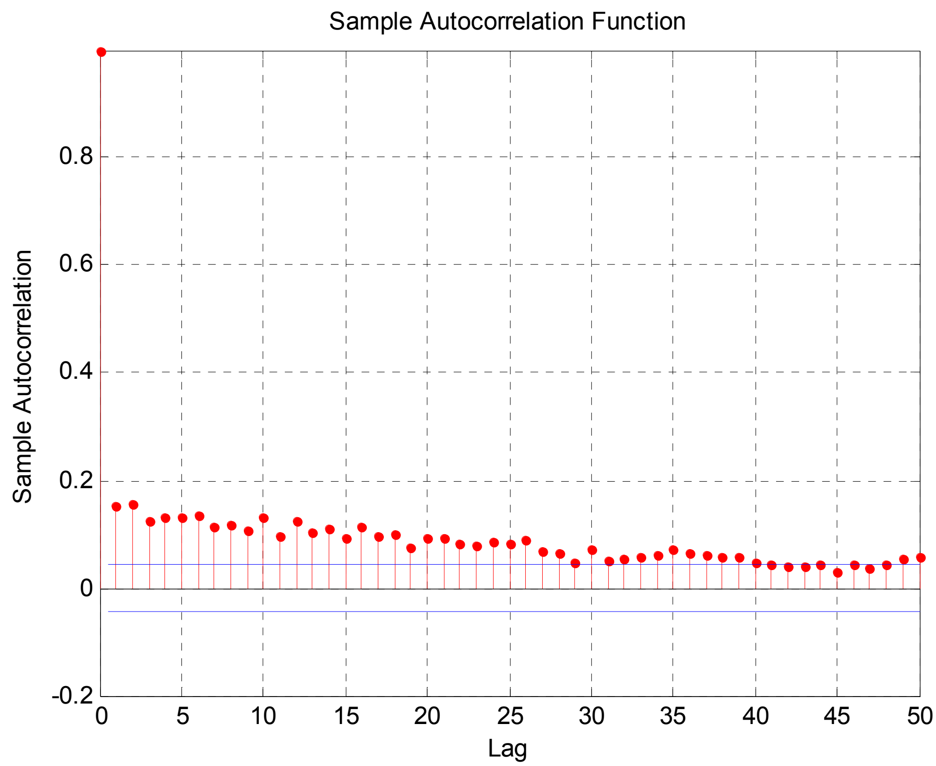

Figure 2.12: An individual example of the results of an Autocorrelation Function. While the correlation at Lag 0 is predictably perfect, the correlation tapers off as lags increase. The correlations remain above chance until Lag 29, when the value is not far enough above the line.

can repeat this for as many lags as we would like, given the timeseries has enough samples. As seen in Figure 2.13, after an initial perfect correlation at Lag 0, the correlation of each subsequent lag decreases. These series of correlations over lags is referred to as the Autocorrelation Function (ACF) and is mathematically related to the power spectra fit by Gilden's analysis as the power spectra (which is used to determine slope) is the Fourier transform of the autocorrelation function.

Another way of looking at this data is called the Partial Autocorrelation Function (PACF), which allows us to view autocorrelation in a slightly different way.

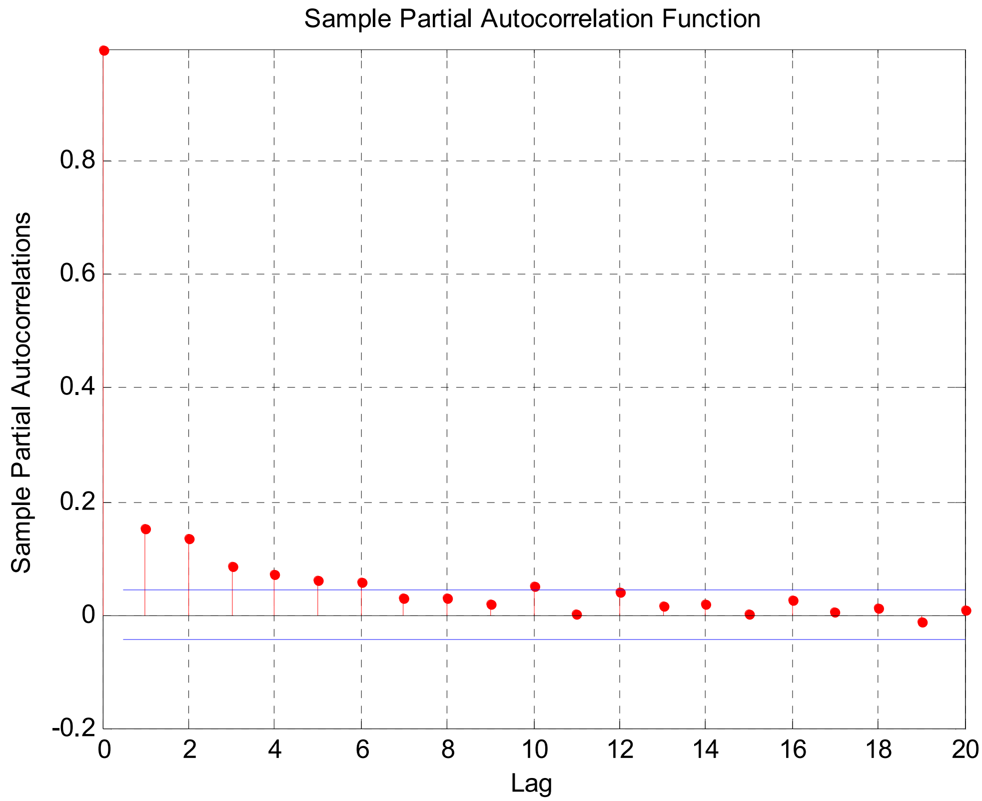

Figure 2.13: An example for a partial autocorrelation function for an individual film. This output is similar to the autocorrelation function in Figure 2.13, but each lag takes into account the influence of previous lags, resulting in a steeper drop in correlation.

The partial correlation function also shows the data's correlation in the form of lags; however, it discounts the influence of the previous lag before making the calculation. In this sense, it is a good way to help decide how many lags may be appropriate to include when modeling the data. As seen in Figure 2.14, the PACF remains significant for fewer lags, indicating that the correlation at later lags are being influenced by the strength of the earlier lags. Although there are no hard and fast rules for determining how many lags are appropriate to include in an autoregressive model, the Partial Autocorrelation typically shows a steep descent in correlation that can be cut off once the correlation falls below some statistical parameter.

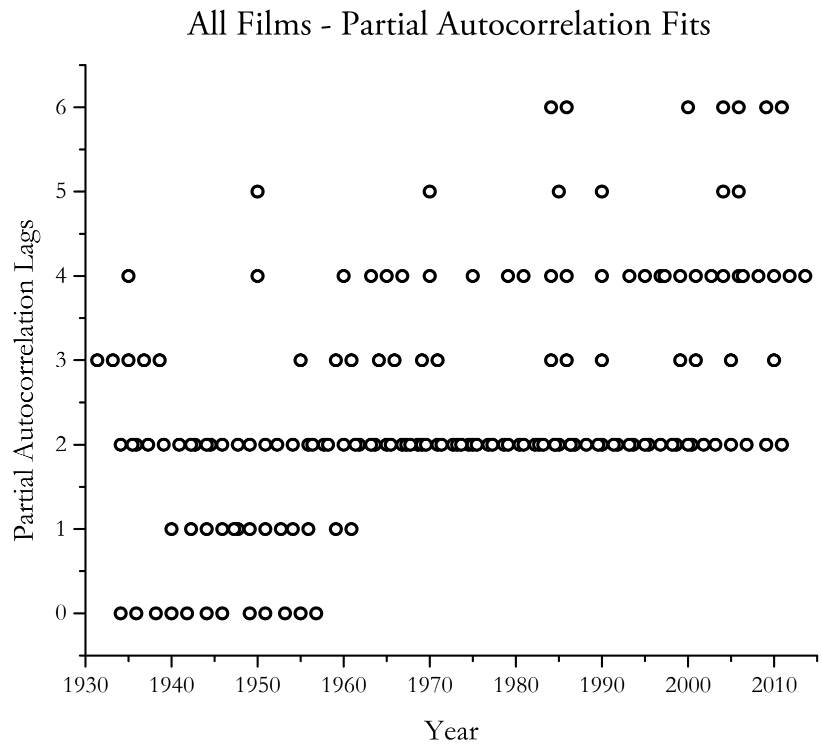

Figure 2.14: Films from the 1930's to 1960's have less significant lags in their partial autocorrelation function, suggesting more long-range correlation in contemporary film.

Another way of determining how many lags to include in the model is to fit a series of autoregressive models with increasing parameters and test whether predictive power remains significantly higher than chance. We used this method (combined with a smoothing function for the Partial Autocorrelation function) when first fitting our data (Cutting et al., 2010). This was also used on updated data in an online response posted to the Cinemetrics website by James Cutting (2014).

Analyzing the shots in our database yields remarkably similar results when fitting serial autoregressive models or utilizing the PACF. Earlier films exhibit no reliable correlation past one or two shots. Later films are usually significant past one or two shots. The positive trend by year is strong, reliable, and predicts that the

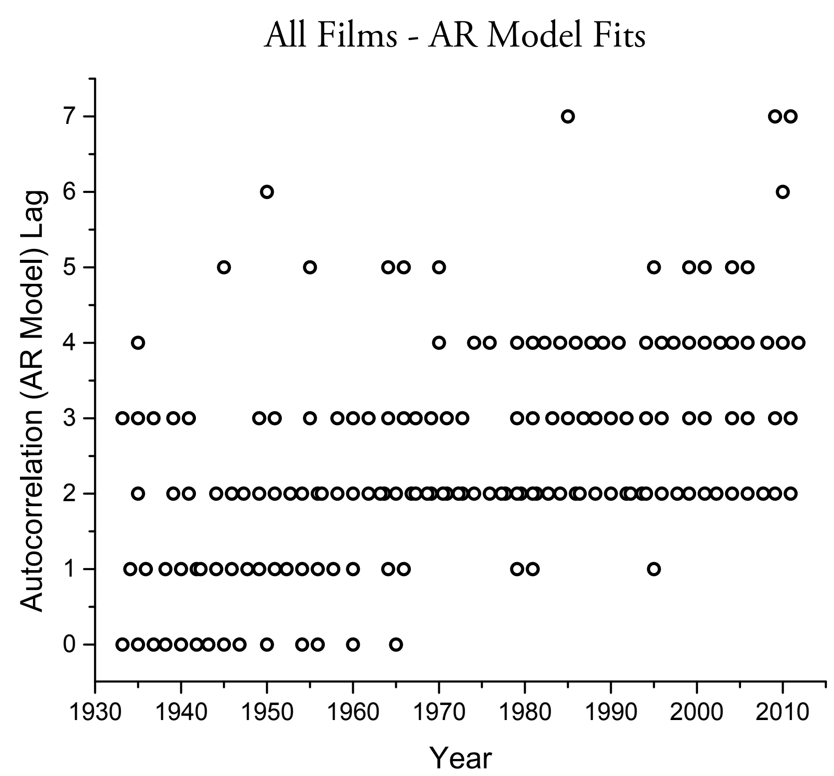

Figure 2.15: Fitting films with the autoregressive model produces similar results to the partial autocorrelation function; however, there are lags that are far more significant.

correlation between shots increases by one roughly every thirty years. This is definitely a slow change, but is reliable throughout our sample.

What does it mean for a film to contain shots with correlations at higher lags? In one sense, these films are clearly not random as each significant lag is evidence of divergence from the expected white noise values. Having a larger significant lag suggest a departure from white noise at an increasingly large scale, suggesting that shots of similar length are clustering together throughout the film. Whether this type of departure is a better description for a Brownian or 1/*f* pattern is difficult to tell, as

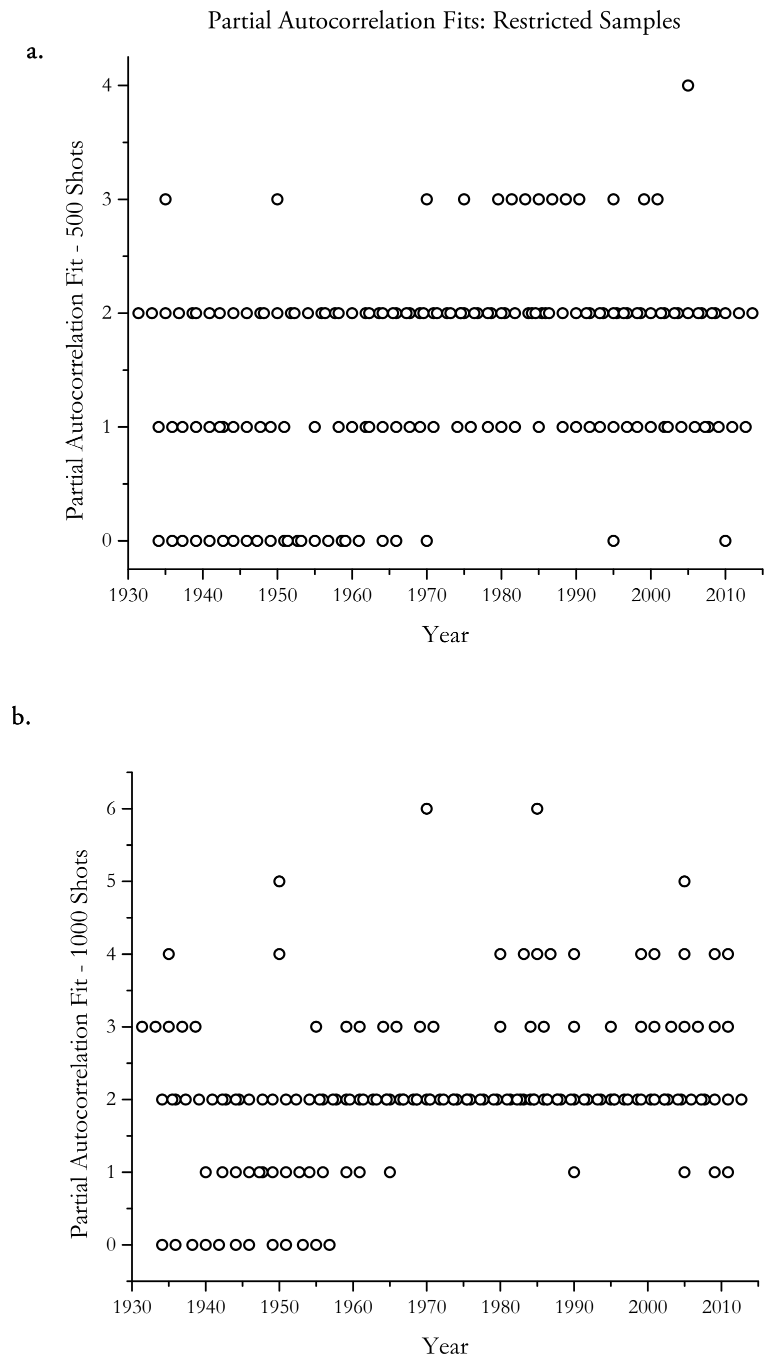

Figure 2.16: This figure shows that truncating films down to 500 shots eliminates the higher-lag fits, as seen in the top graph. This is not particularly surprising, as decreasing samples will often diminish effects, but highlights that film length is impacting significant lags. As the lower graph shows, when films are truncated to 1,000 shots these higher-lag values reappear, suggesting that the number of shots is influencing model fits.

both are departures from white noise. In essence, a 1/*f* pattern needs to have correlations at later lags, but correlations at later lags do not immediately suggest a 1/*f* pattern.

Another valid issue with looking at the ACF and PACF is that films with fewer shots may not be significant at higher lags simply because of fewer shots in a sample. Given that earlier films tend to have shorter shots, it is possible that the increase in our database could be driven by the increasing number of samples for each film which in turn allows for films to become significant at higher lags. A simple way to check the effect of this is to truncate the longer films and re-run the analysis. Rather than try to resample the films and introduce any kind of smoothing artifacts, 500 shot chunks were removed from the middle of the film and analyzed. As shown in Figure 2.17a, keeping all films trimmed at or below 500 shots was enough to diminish, but not eliminate the trend of increasingly significant lag over time. The most noticeable change is that while previous fits were significant out to five or six lags, the range with the trimmed films has been truncated. Figure 2.16b shows that by increasing the maximum amount of shots to 1,000 we can see that the longer lags become significant.

The Hurst Exponent

After World War II Harold Edwin Hurst had a problem. As a British civil servant who had expatriated to Egypt, he wanted to find a way to better harness the Nile River's flooding to fill reservoirs that would provide a water supply during dry months as well as stop massive flooding. The question of the capacity of the reservoirs is a difficult one because the Nile's high water level often changes, and while seasons play a major role, there is variation on the weekly, monthly, and daily scales that can be difficult to predict. In an effort to solve, the problem Hurst invented a method in which he could use historical data to model the range of variation in the Nile at different timescales. This method is fairly straightforward and consists of selecting a set number of values in a timeseries, dividing the range divided by the standard deviation of that series, and then repeatedly plotting those values on a log-log plot (Hurst, Black, & Simaika, 1965; Hurst, 1951). Finally, a linear fit should reveal an increasing amount of variation over time, as seen in Figure 12.18. This method took time to be understood and adopted by statisticians who, at the time, continued to recommend autoregressive models (Seneta, 2010). It took until 1965 for mathematician Benoit Mandelbrot to understand the importance of Hurst's discovery, promote it in the mainstream, and name the process in his honor (1965).

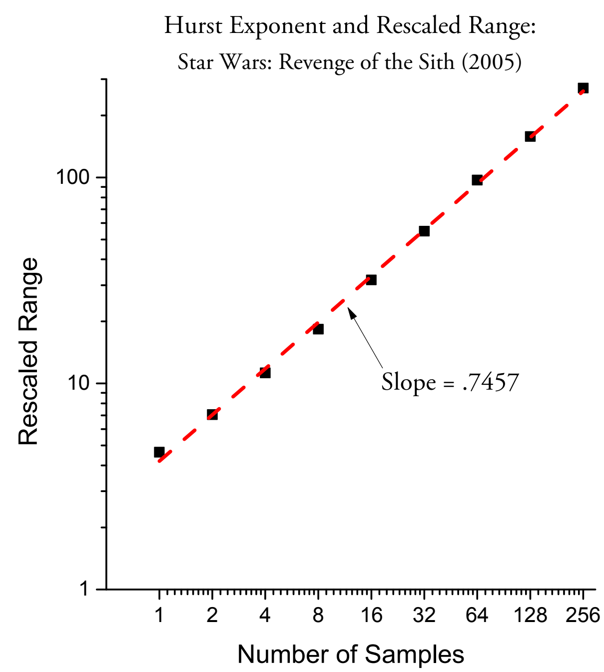

Figure 2.17: As the as the number of samples increases for long-memory systems, the range of possible values also increases. In this example, the slope of the increase is fit linearly on a log-log plot.

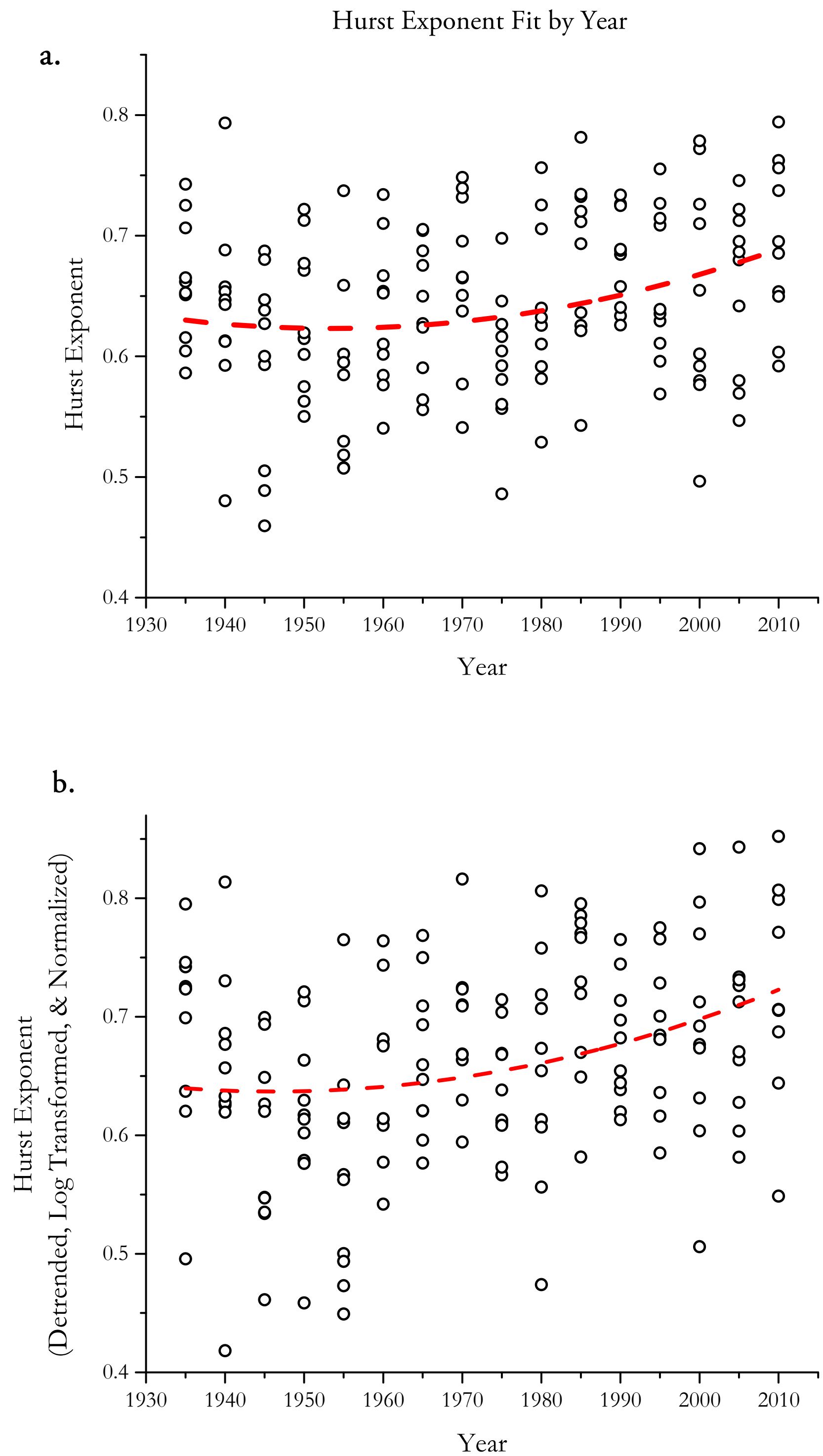

Figure 2.18: Fitting each film using Hurst's method suggest a more dramatic movement towards 1/*f* structure (which is a value of 1 on the Hurst Exponent scale) than using Gilden's method, as seen in the top graph. The lower graph shows the same results, but after each film had been pre-processed with modern techniques such as detrending, log-transforming, and normalizing each film.

This method discriminates between white noise and a 1/*f* pattern by determining whether the process has a 'long memory', another way of stating that longer lags have influence. For Hurst, this meant that a particularly intense flooding season may have long-reaching effects that can influence the next season and beyond. This measure has some issues detecting the difference between 1/*f* noise and Brownian noise. Hurst dimensionality is on a scale from zero to one. A timeseries that does not exhibit a long memory will produce an exponent of .5. As the value of the exponent increases, so does the memory of the system. Much like with the longer lags in the ACF and PACF, it is necessary that a 1/*f* pattern has a long memory. But simply observing long memory characteristics is not sufficient to identify a process as 1/*f*.

After analyzing our dataset using Hurst's method we can see that for our dataset, films exhibit a larger Hurst exponent throughout the decades, as seen in Figure 12.19a. A major concern with Hurst's method is that non-stationary and non-gaussian timeseries data may provide false positive exponent fits. Figure 12.19b shows results of the same method, but using data that has been modified before analysis by detrending, log-transforming, and normalizing shot lengths.

Detrended Fluctuation Analysis

Many years after Hurst's method was published, scholars began looking for a type of rescaled range analysis that could be used in non-stationary situations, like the automated analysis of abnormal heart rhythms and the detection of nonrandom DNA sequences in the genome. One of the best-known methods is called Detrended Fluctuation Analysis, which purports to detect long-range memory in timeseries

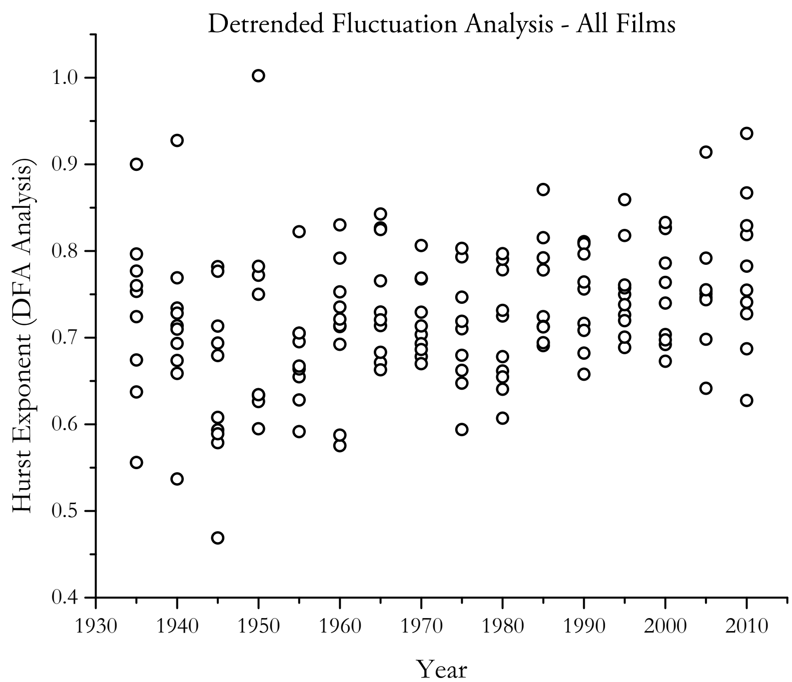

Figure 2.19: The results of running each film through detrended fluctuation analysis is very similar to the results of vanilla Hurst dimensionality, however the films have less variation per decade and show a much more straightforward increase over time.

that are non-stationary (Peng, Havlin, Stanley, & Goldberger, 1995). Despite widespread adoption and acceptance, this method does not appear to detect long-range dependencies any more accurately than using Hurst's original method, at least while analyzing data with 20,000 samples (Bryce & Sprague, 2012).

Figure 12.20 shows the results of our shot length data modeled with DFA. The output of the model is the Hurst Exponent, so direct comparison between the DFA and rescaled range analysis are possible. The data show the same trend of slowly-increasing dimensional fit throughout the decades, but the films within each year appear to have less variability than using the classic version of rescaled range analysis.

ARMA and ARFIMA

After Gilden published his work, attempts to replicate and interpret the findings lead to questions about whether his power-spectra fitting analysis was the best way to test for 1/*f* and/or long-range dependency. One notable set of papers collected data in an attempt to see whether the noise generated within reaction times really has long-range dependent properties or if the 1/*f* pattern observed may simply be due to short-range dependent processes (Farrell, Wagenmakers, & Ratcliff, 2006; Wagenmakers, Farrell, & Ratcliff, 2004). The question of 1/*f* fitting doesn't merely apply only to Gilden's work, but also work published before Gilden's results (Gottschalk, Bauer, & Whybrow, 1995) as well as recent findings in social psychology which report that reaction times of subjects attempting to modulate their responses in a racial bias task exhibited less 1/*f* noise (Correll, 2008, 2011). It is worth noting that the results from the racial bias task have been shown as not significantly replicating, even with two controlled and high-powered attempts (Madurski & LeBel, 2014).

In order to test whether long-range memory or short-range noise are responsible for a timeseries to be misclassified, Wagenmakers, Farrell, and Ratcliff suggested a methodology that should be able to separately model long and short range dependencies (Wagenmakers et al., 2004). The first step of this is to fit an Autoregressive Moving Average (ARMA) model to the timeseries and calculate the Akaike information criterion (AIC), which is a method of determining how well a model fits the data (Akaike, 1980). The fit for the short range model is compared to the AIC for a new Autoregressive Fractionally Integrated Moving Average

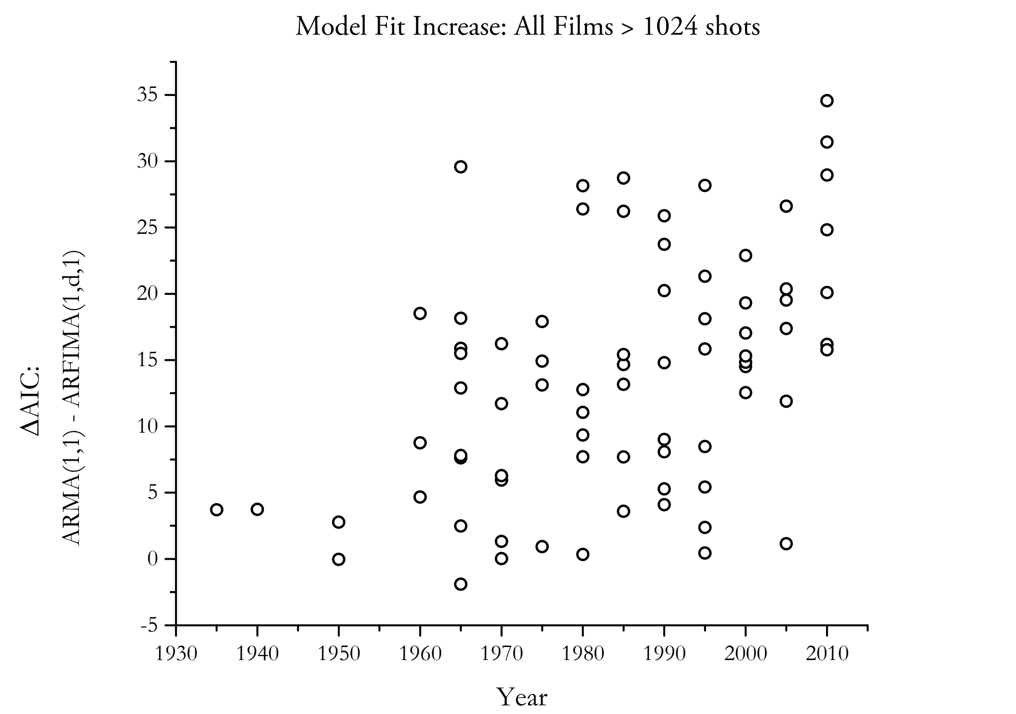

Figure 2.20: For films with more than 1024 shots, many films in the sample are better described with an ARFIMA model than an ARMA model, suggesting that these films can be appropriately described as evidence for a long-memory process.

(ARFIMA) model that is fit to the data. If the ARFIMA model does not produce a much better fit than the simpler ARMA model, then it makes sense that the timeseries does not contain anything other than white noise. The code necessary to perform this analysis has been provided through Farrell's website, with the caveat that all timeseries must have at least 1024 samples (Farrell, 2008).

Whittle Estimator

Peter Whittle left New Zealand in 1950 to get his PhD at the University of Uppsala with noted economist and statistician Herman Wold. At the time, Whittle didn't speak a word of Swedish, but learned the language over six months by translating a book on the geostatistics of forestry (Kelly, 1994). Whittle's graduate

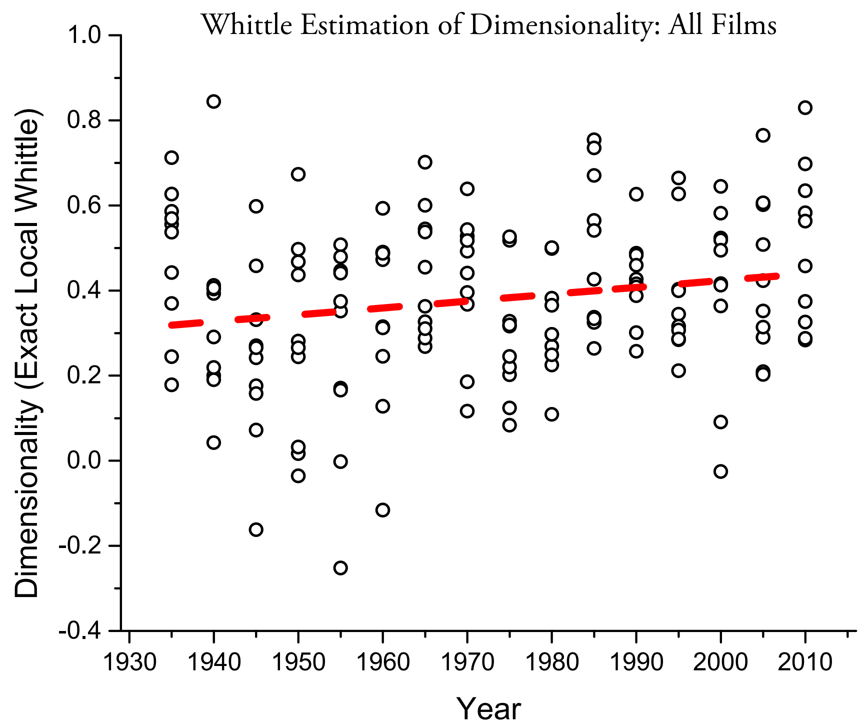

Figure 2.21: The Whittle estimator also shows our data as slowly becoming more 1/*f* over time. It is worth noting that some of the earlier films are modeled as being slightly anticorrelated, with a dimensionality value that is less than zero.

work expanded that of his famous adviser by innovating new methods to deconstruct *any* timeseries into different components that describe the structure of the change in the data. (Whittle, 1953). It took many years for this mathematical trick to be used in a functional way (Kuensch, 1987) and even longer to be implemented into the statistical process that bears Whittle's name today (Shimotsu & Phillips, 2000).

An exact description of the Whittle estimator is largely beyond the scope of this dissertation; however, an attempt at a short description follows. The Whittle estimator is, in essence, a combination of both the autoregressive and power spectra based methods described before. At the heart of the algorithm is the same type of classic log-log fit, however the points of data that are fit by the algorithm are influenced by autoregressive fits. This influence is based upon the expected set of covariance properties originally calculated by Whittle, and actually change the estimation of each point based upon the parameters expected at individual lags given different levels of correlation. This then helps to limit the influence of short-range correlations and noise in a principled manner, rather than relying on fitting the data by brute force or simply excluding noisy data altogether.

Comparison of techniques

In order to accurately review the performance of each dimensionality-estimation technique, we will need to generate sample datasets that have a known dimensional component. The basic concept of how to generate the self-similar noise required to do the job was outlined a long time ago (Mandelbrot & Van Ness, 1968); however, modern techniques create the random set in the frequency domain, rather than attempting to simulate complex temporal patterns such as Levy flights or other Markov processes in the time domain. For this dissertation, a sample timeseries will be randomly generated by generating power spectra with a known slope, combining the spectra with randomized phase elements, and then constructing the timeseries using an inverse Fourier transform.

Each technique was tested using randomly generated timeseries of different lengths (100, 250, 500, 750, 1000, 1500, 2000) as well as with a range of dimensionality values from white to Brownian noise (-0, -.25, -.5, -.75, -1, -1.25, -1.5, -1.75, -2). Each combination of categories was simulated 5,000 times resulting in a total of 315,000 simulated timeseries that would be modeled by each technique. This allows for the mapping and direct comparison of different methods that do not output precisely the same range of values, but still may have some ability to discriminate between each dataset's dimensionality. Each technique was judged using two major criteria: the ability to identify fractal dimension as well as remaining consistent with different numbers of trials.

Discrimination of dimensionality at 2,000 samples

For our first evaluation, we will look at how well these techniques work under the ideal condition of 2,000 datapoints. By holding the amount of trials constant, we can get a good look at what to expect from the output of each method. As seen in figure 12.23a, given 2,000 datapoints and 5,000 iterations, the power-spectrum and rescaled-range type analyses do an excellent job of measuring dimensionality. With the exception of the classic Hurst method, these methods appear to have a linear response to the increase in dimensionality, albeit through different ranges. Gilden's analysis ranges from 0 to 2, the Whittle Estimator is ranged from 0 to 1, the Hurst estimator ranges from .5 to roughly .9, and DFA runs from .5 to 1.5. In ideal conditions, it looks like our version of Gilden's technique, DFA, and the Whittle estimator are all good methods.

A more curious pattern emerges when looking at the output for the autoregressive methods in Figure 12.23b. While the PACF does increase while heading towards a 1/*f* pattern it heads back towards Lag 1 in a parabolic pattern. This makes sense – Brownian noise features stronger close-range relationships. When looking at the AR-refit method, the results are a little more difficult to interpret. It looks as if lag increases when encountering Alpha values between 0 to -.7, but it

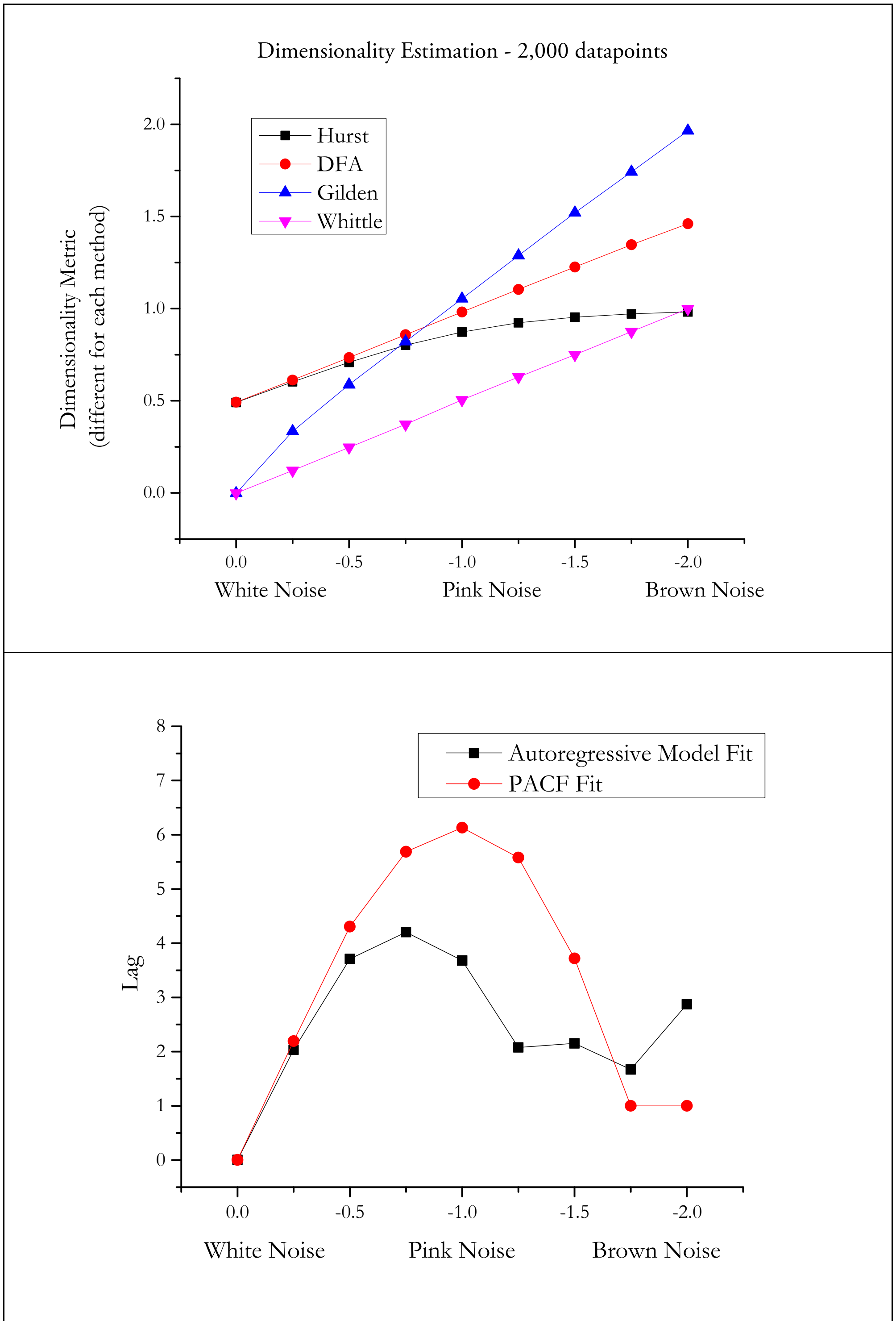

Figure 2.22: Results from simulated datasets show how each method compares in ideal conditions of 2,000 samples. Across both figures, Gilden's analysis, the Whittle estimator, and Detrended Fluctuation Analysis do a good job of linearly fitting each type of noise. Both autoregressive models show an inability to distinguish between white and brown noise.

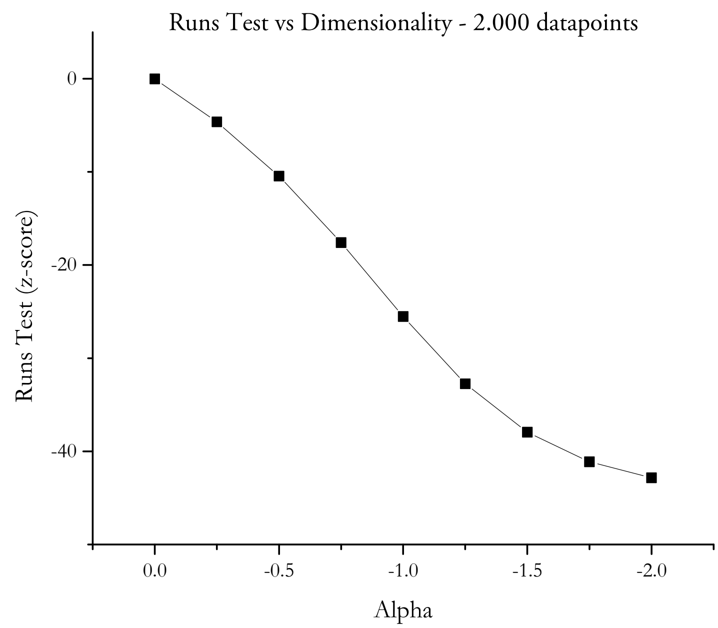

Figure 2.23: Despite being one of the older and simpler method, the z-score from the Wald-Wolfowitz runs test does a good job modeling differences between white and pink noise.

also doubles back towards the lower lags as alpha increases only to rise again once reaching Brownian noise.

The final method is the Wald-Wolfowitz Runs test, which does a remarkably good job of discriminating between random and 1/*f* noise but appears to tail off between 1/*f* and Brownian noise. Truly random noise is still classified as random through hypothesis testing, however datasets with *any* amount of non-randomness are all considered statistically different from chance, so to compare them we will look at the z-scores. A similar comparison is simply to look at the ratio of actual runs vs. expected runs, a comparison which yields similar results.

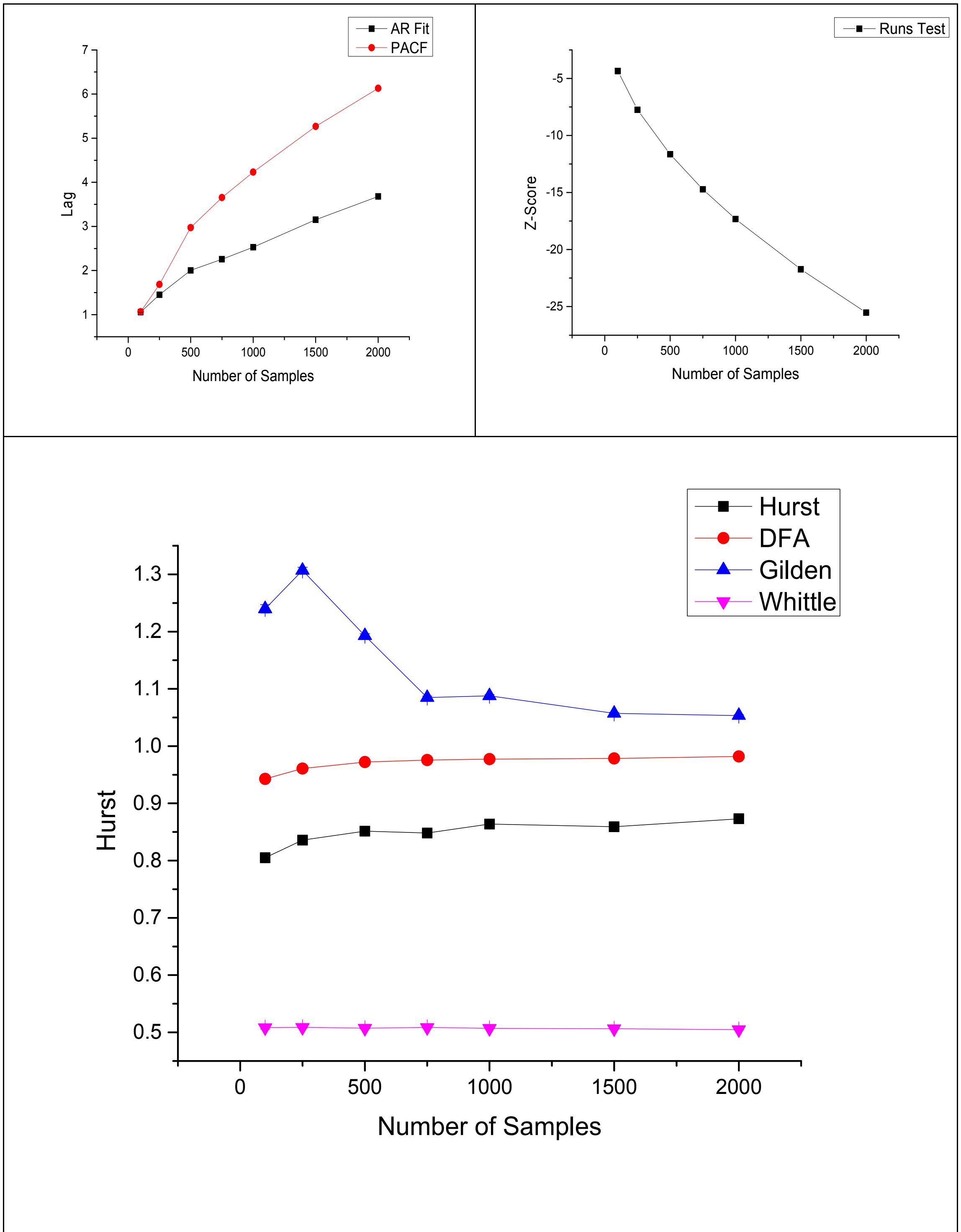

Figure 2.24: A series of graphs show how sensitive each method is to sample size. Ideally, a method would continue to report the same value for a $1/f$ pattern with 100 or 1,000 samples and would show up as a flat line on the graph. The Hurst and Detrended Fluctuation Analysis remain relatively flat; however, the Whittle estimator does the best job of remaining consistent.

Consistency across samples

Another concern with fitting the dimensionality of films is that the number of shots are simply not consistent and range from roughly 200 to 2,000 shots. In order to observe how each of these methods works with different numbers of samples, we can keep the type of noise constant but timeseries of different lengths. It is possible to make comparisons for all types of noise, but in the interest of simplicity, I chose to compare 1/*f* patterns across different sample lengths.

As can be seen in Figure 12.25a, the autoregressive methods have a tendency to increase number of significant lags as the timeseries becomes longer. This can be attributed to the fact that, although statistically the ACF and PACF functions may look the same, the number of samples influences the statistical tests used to decide whether a lag is significant. A similar story can be told for the Wald-Wolfowitz runs test in Figure 12.25b, where, unsurprisingly, the z-score is more dependent on the number of samples in the dataset than any underlying dimensionality. While these methods could be used to compare against samples of similar length, they do not appear to be deal well with the films of different length in our database.

Figure 12.25c shows the relationship of the periodogram-based (Gilden and Whittle) and rescaled-range (Hurst and DFA) approaches. Overall, the rescaled-range approaches slightly underestimate the dimensionality of the dataset until reaching roughly 1,000 samples. Our version of the Gilden analysis displays the opposite pattern, systematically *overestimating* the dimensionality of timeseries with less than 750 samples. Whether this is due to the method itself or simply due to changing how many parameters are used to fit the model, an overestimation for films under 750 shots has fairly large ramifications for how we talk about change in film editing over time. With more bias in shorter films, this would make the move towards a 1/*f* pattern *more pronounced* over time given that shorter films load the earlier years of our sample.

Despite these considerations, the most consistent method is the Whittle estimator, which is balanced enough (over 5,000 trials) to identify 1/*f* noise whether the dataset was composed of 100 samples or 2,000 samples. On average, the Whittle estimator not only accurately categorizes the dimensionality of the timeseries (Figure 12.23b) but also remains consistent across samples of different sizes, and is clearly the best method for the simulated data and probably for our film database as well. The question remains: how consistent is it on a sample-by-sample basis?

As the Whittle estimator accurately measures dimensionality, it is possible to use a simple statistic-like standard deviation (rather than Mean Squared Error) to evaluate how accurate the model is from timeseries to timeseries Figure 12.26 shows the expected dropoff in error given the increase in trials. This value is influenced both by the Whittle

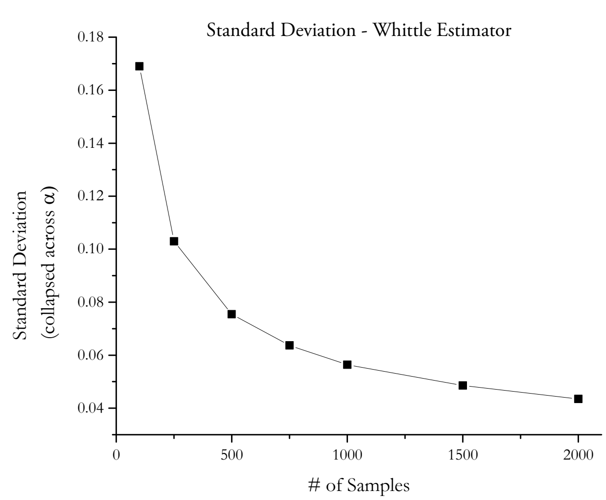

Figure 2.25: It is unsurprising that with more data, the Whittle Estimator gives more accurate results, however the decrease in error is not linear.

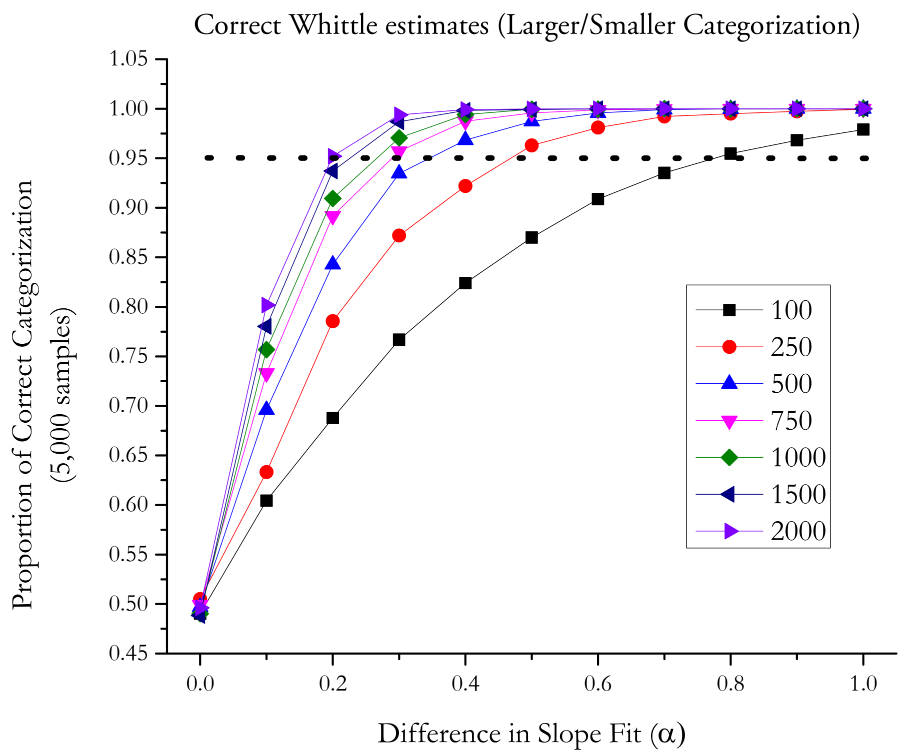

Figure 2.26: While mapping the performance of the Whittle Estimator with simulated data, it is clear that any difference in α less than .2 cannot be reliably detected even in ideal conditions. If the difference in α is greater than .3, it is not necessary to have more than 750 samples to achieve a better than 95% correct categorization rate.

estimator's ability to discriminate as well as by the noise within the process of creating a random timeseries, which will not always be perfectly representative. The error for the Whittle estimator drops off disproportionately as more samples are added, with the largest decrease being between 100 and 250 samples.

Another way to look at the Whittle estimator's accuracy is to test the tradeoff between differences in dimensionality and number of samples when discriminating between two sample timeseries. In order to simulate this, two random timeseries were generated to have either no difference in dimensionality (i.e. two random sequences) or some amount of dimensionality (i.e. comparing a random sequence with an alpha value of zero to one that is slightly nonrandom and has a value of -.1). If the Whittle estimator was able to discriminate, which value was higher or lower, the comparison was considered a successful categorization. This comparison was carried out for ten slope values between white and pink noise (0 through -1) for a variety of different timeseries lengths (100, 250, 500, 750, 1,000, 1,500, 2,000) with 5,000 iterations.

As seen in Figure 2.27, the Whittle estimator was able to accurately detect large differences (such as that between white and pink noise) over 95% of the time even when the number of samples was as low as 100. Smaller differences in slope required more samples to detect. When detecting the difference between α values less than .5 (or halfway between white and pink noise) it was adequate to have only 250 samples.

Ramifications and Interpretation

Understanding the characteristics and limits of our different dimensionality estimation techniques can give us better insight into what we can detect within a particular set of data. Within our results, while the use of our version of the Gilden technique and Autoregressive modeling may not have been the most optimal method of analysis, the interpretation of the data does not necessarily change. Hollywood film appears to have been gradually and modestly moving toward a 1/*f* pattern in their shot lengths. We believe that our sample is representative of the overall trends in film, but we cannot definitively conclude anything without more samples than collected.

Given the use of the Whittle estimator instead of other methods, how fast are films changing? Using a linear fit to model the change over time results in a less than impressive R^2 value of .031. The slope appears even less impressive, with the dimensionality measure (in α) increasing at a rate of ~.003 per year, though it is significantly different from zero. Given the rate of increase and the starting point in 1935, we could project that if the trend continues linearly it will take until 2147 for the average film to show $1/f$ structure. While this may seem like a long time away, in the broader context of how language and art changes over time a couple hundred years is a relatively quick change.

Do we even have the power to detect this small of a change? The comparisons in Figure 2.27 can help to show where we stand. Detecting the change between each five-year increment cannot be expected to work if we take the linear fit seriously. Fitting the data with a parabolic fit is actually much better, but in the interest of providing a conservative estimate, we will proceed with the linear fit. The expected increase of .015 over five years will only be detected slightly more often than chance, and the increase over ten is not much better. Once we get to the increase over the full range of our dataset (α increases roughly .225 over 75 years) we can observe reliable changes. With the average film in our database containing 1,179 shots, we can see that the comparison between two sequences with a difference in α of .2 will be accurately categorized 90% of the time. Taking into account the multiple comparisons within our dataset, we can conclude that our database and analysis has a reasonable chance to find even the small changes we observe in film.

Will film stop once it reaches a 1/*f* pattern or continue towards a random walk? It is impossible to say and possibly foolish to speculate; however, we will not really know any disconfirming evidence until so far in the future that this dissertation will be long forgotten. It is this author's opinion that the 1/*f* pattern is special and appealing to humans. It is not clear why this is the case, but I would not be surprised if films edge towards 1/*f* and then turn sharply back towards randomness due to some kind of trend or change in filmmaking. We can predict that films will change linearly all we want, but the only constant in art is (and should be) change and upheaval of the status quo. If I am a retiree and films continue to change and evolve in the same way they have over the past hundred years, I will be *very* disappointed.

Is it appropriate to fit a power law to these films in the first place? As the new methods and observations of physics and mathematics come into vogue, it is understandable that these new methods will inspire many to jump on the bandwagon of power laws. Our work with film is not exempt from this characterization, utilizing methods developed by others to characterize datasets in new ways that aren't fully realized or understood. Most methods of determining dimensionality used in this dissertation were developed to detect pure power laws. Although these systems appear to work well on randomly generated data, do our films themselves actually look like power laws?

Individual films definitely do not follow a strict power law. This makes sense, as films have a great deal of variation and are not random samples of some underlying process. As seen in figure 2.28, the 1935 film Anna Karenina does not follow a strict

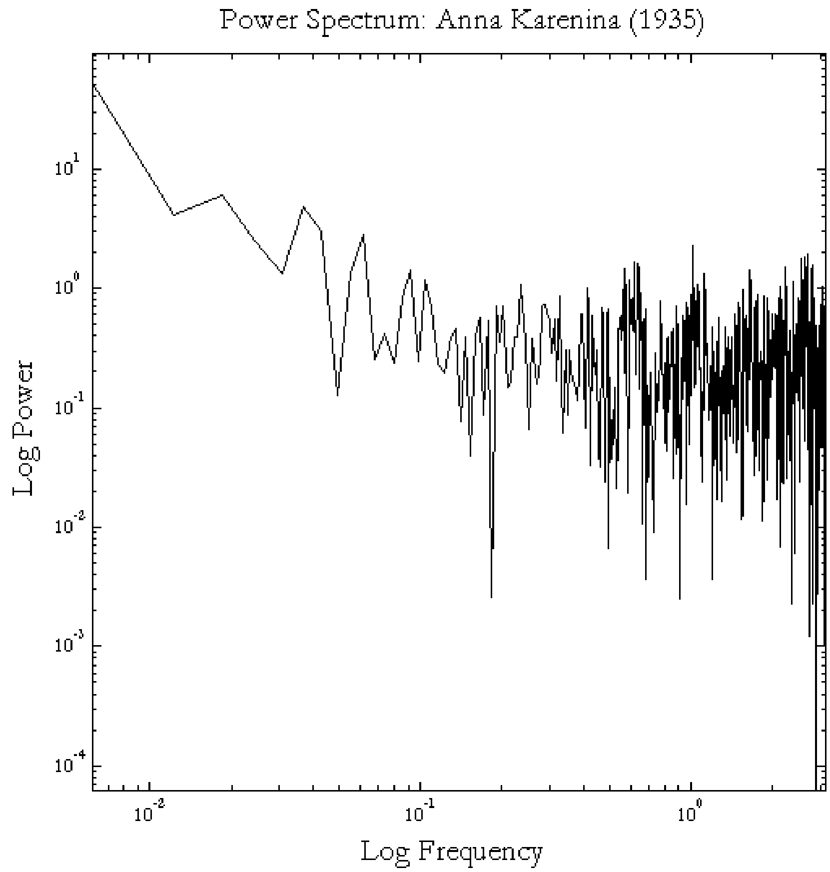

Figure 2.27: It is clear that this film does not follow a strict power law, as all frequency components do not fit on a straight line on this log-log plot.

power law. It is possible, however, to average all films' power spectra together in order to get an idea of whether or not films in general look similar to the power spectra that are generated by power laws.

Figure 2.29 shows the averaged power spectra for all 160 films in our database. The power for each film was normalized by dividing each film's response by the area underneath the curve. The averaged power spectra do not look perfectly in-line with what would be expected from a pure power-law, which would show up as a straight line.

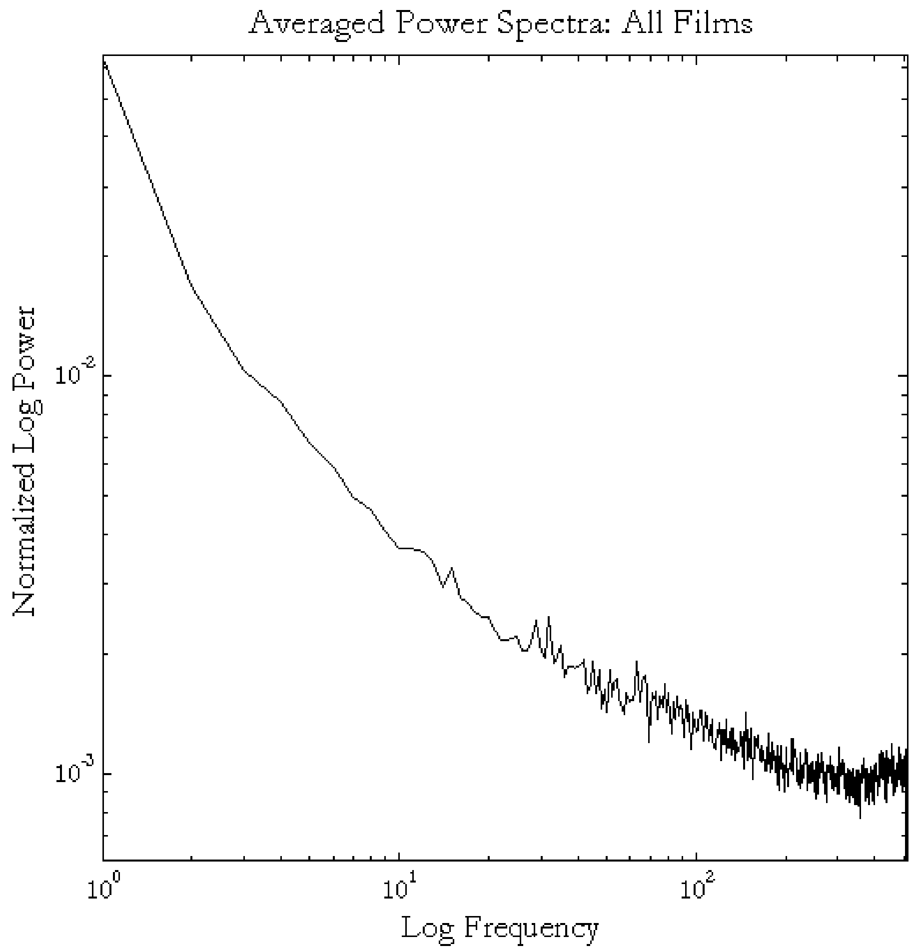

Figure 2.28: A log-log plot of normalized power at a variety of different frequencies averaged across all 160 films in the database.

Given this divergence from a pure power law, is it incorrect to proceed with fitting dimensionality parameters to the data? The answer to this question has actually been addressed by physicists who describe certain systems that may be described as some kind of pink noise with white noise added, which changes the properties of the spectrum (Pilgram & Kaplan, 1998). Gilden also knew of this issue, and attempted to correct it by fitting both pink *and* white noise in his model (2001). Other worried that only fitting certain data points may lead to overestimation of slope (Wagenmakers et al., 2004). Regardless, while it appears that Hollywood film doesn't perfectly follow a power law, it should be possible to find ways to characterize different types of noise within each timeseries.

Conclusions

Over time, films are changing in a variety of ways, one of which being the increase in the display of a 1/*f* pattern. You can measure how well a timeseries fits the properties of a random, 1/*f*, or Brownian pattern using a several different methods. For our purposes, the Whittle estimator appeared to be the best performer by remaining consistent even with smaller datasets. We have proposed in previous work that the fact film is moving towards the endogenous patterns found in human attention may be more than just a coincidence (Cutting et al., 2010). While there are many things that make a film enjoyable, it's an interesting thought that editing patterns may contribute to a film's ability to hold our attention.

CHAPTER 3

Historical Data and the 1/*f* Pattern

What is driving changes in film editing patterns?

The finding that films are getting closer to following a 1/*f* pattern over time can be interpreted in a variety of ways. The original interpretation presented in our results is that film is slowly evolving to match the patterns found within endogenous patterns of attention (Cutting et al., 2010). Subsequent discussions and critiques of the data revealed another intriguing possibility – maybe our brains have been changing radically over the past 75 years and film is simply mirroring that change either in real-time or with a slight lag.

It is not impossible that such a large scale shift in cognition could escape the attention of psychologists. In 1984, Political Scientist James Flynn published a paper that reviewed IQ scores that had been collected since 1932. The calculation of IQ scores has been standardized to the population taking the test so that IQ was really just a percentile measure. Flynn looked back on previous averages and realized a startling fact –Americans have shown steadily increasing IQ scores over time (1984). This trend would have been glaringly apparent to any who had taken the time to look for it with even a decade or so of data, but it took until the 1980's for someone to challenge the assumption that fluctuations were systemic rather than the results of random variation in the data. In fact, it was noted that IQ scores were rising a decade before Flynn published his paper, but the breadth and ubiquity of the phenomena was not pursued beyond mentioning the trend. (Thorndike, 1975). Furthermore, this result isn't just found in America but in other countries around the world (Flynn, 1987).

The finding that people were getting more intelligent (at least on paper) was a rare bit of good news from the sciences, and the quest to claim credit was undertaken from various disciplines. One of the best books to cover this topic featured arguments from Flynn as well as experts from other fields (U. Neisser, 1998). Flynn posited that factors such as education and urbanization might account for the effect, with respective fields bolstering the importance of their field's contribution. More recent takes on the Flynn effect typically emerge from individuals championing the importance of their own fields. This has included claims that greater genetic diversity (read, less inbreeding) has facilitated an increase in intelligence (Mingroni, 2007), that fertility patterns shape early learning enough to contribute to the effect (Sundet, Borren, & Tambs, 2008), and that increasingly complex visual media increase our visual processing abilities (Barber, 2006).

Might an increase in observable $1/f$ patterns in film be correlated with the Flynn effect? Possibly, but in order to comfortably infer some kind of connection we need to see if $1/f$ has been changing over time, and preferably collect some samples before the advent of film. Unfortunately, the discovery of the Flynn effect was only made possible by systematically collecting several decades of comparable data across a variety of participants. In our current short-term application driven grant economy, we are not particularly likely to collect any longitudinal data without a very specific hypothesis. Timeseries modeling of reaction time data only began in twenty years ago and has never been a mainstream area of inquiry for psychologists. Any attempt to use modern modeling techniques on historical data will require access to trial-level data rather than the mean values typically reported in journal articles.

Unfortunately, most psychologists are not in the habit of archiving and preserving raw data beyond five or ten years after a study was published. This was especially true in the pre-digital era when a faculty office could quickly fill up with boxes of collected data. Some researchers are even forced to destroy older data in accordance with privacy policies. While it might seem that digital data would be more likely to be saved, issues with formatting, changing file formats, and obsolete media storage make records collected before the advent of the internet also difficult to access. Luckily, the Center for the History of Psychology Archive at the University of Akron exists to try to preserve notable and rare artifacts including marquee items such as Bandura’s Bobo doll, Milgram’s fake shock machine, and a prototype Skinner box. They also have archives containing data, stimuli, personal correspondence and 50,000 rare books previously owned and written by notable psychologists.

For this project, three raw sets of data were collected from the Akron archive on November 10th and 11th of 2011. The three datasets were found with assistance from various members of the staff, but Senior Archives Associate Lizette Royer was instrumental in tracking down the data. Materials were often contained in bound notebooks, so they were laid flat and photographed using an SLR camera, converted to PDF, and transcribed by hand. What follows will be an individual description of each dataset and a quick analysis of whether the dataset is strong enough to help make a conclusion, followed by a post-mortem of the general results and commentary.

Walter R. Miles, Paul Brainard, and Hugh M. Bell

This dataset is taken from the private papers of Walter R. Miles, a prominent psychologist who worked in a variety of areas in experimental psychology[1]. Miles became interested in psychology when he joined the lab of a Cornell doctoral candidate who was writing his dissertation and teaching at Earlham College in Richmond, Indiana, bringing along books by E. B. Titchener. Miles decided on higher education after he heard of the success that his cousin, Herbert Hoover, had found after graduating in Stanford's inaugural class, even after failing nearly every entrance exam. After positions at Wesleyan, Yale, and later Stanford, Miles took a grand tour of the laboratories of Europe to touch base after the World War I disrupted communications, and came back with an excitement for psychological measurement (Miles, Goodwin, & Royer, 2010). In 1928, he was teaching a course in Advanced Experimental Psychology and assigned his students to create a hands-on lab report using the different devices. It is unclear why Miles preserved this particular lab report, but it travelled with his belongings to locations as far as Istanbul. The students who put the report together were Paul Brainard, who would go on to work for the Psychological Corporation, and Hugh Bell, who would do pioneering work in

[1] Miles' style of experimentation was more or less atheoretical in nature, focusing more on new methodology rather than furthering any particular viewpoint. His flowed from whatever he found most interesting at the moment, and ranged from perfect pitch to breathing patterns in yoga, from alcohol tolerances to distance judgement, and from eyetracking to running blind children through mazes. One of his biographers described his work as "characterized by a tremendous interest and curiosity about many topics, whether narrow or broad in scope, basic or applied in their significance. As he attacked a fresh problem, his work always showed great inventiveness and ingenuity, but by the time it reached publication he was likely to be off on something else." (Hilgard, 1985, p. 420) Miles is probably best known for his work during World War II when he discovered that wearing glasses composed of red-tinted, Corning No... 2403 deep-red lenses could allow pilots to continue daily activities but also be dark-adapted to respond to threats instantly (Miles, 1953).

eyetracking and serve as the Dean of Chico State College and eventually have the student union named in his honor.

Brainard and Bell's experiments focused on using different devices that had been developed by William Miles himself. The students designed a study to see whether lowercase or uppercase words were processed faster, and used a device Miles called a voice chronoscope. This device was essentially a tin-can style microphone connected via taut string to a kymograph, which rotates a smoked piece of paper around a cylinder at a given pace as a feather-tipped needle etches the resulting oscillations into the char. Combined with a small shock from a device that starts the trial by dropping a small card on a table, it is possible to calculate reaction time *very* precisely by measuring the physical distance between the onset and first auditory reaction.

Historically, experimenters would physically measure the distance between onset and offset, calculate an average, and then translate that value into average reaction times for each category. Luckily, Brainard and Bell included the original data as well as kymograph sheets that were typically thrown away. In order for the readings to stay intact, the sheet had to be covered in some kind of fixing solution, which was unfortunately flammable. A box of kymograph paper, comprised of char, wax, and flammable material, was not typically not kept around.

While the kymograph was a very accurate device, each trial session was, unfortunately, limited to the size of the kymograph paper. Brainard and Bell used paper roughly 6.5 inches tall and could record roughly thirty trials before running out

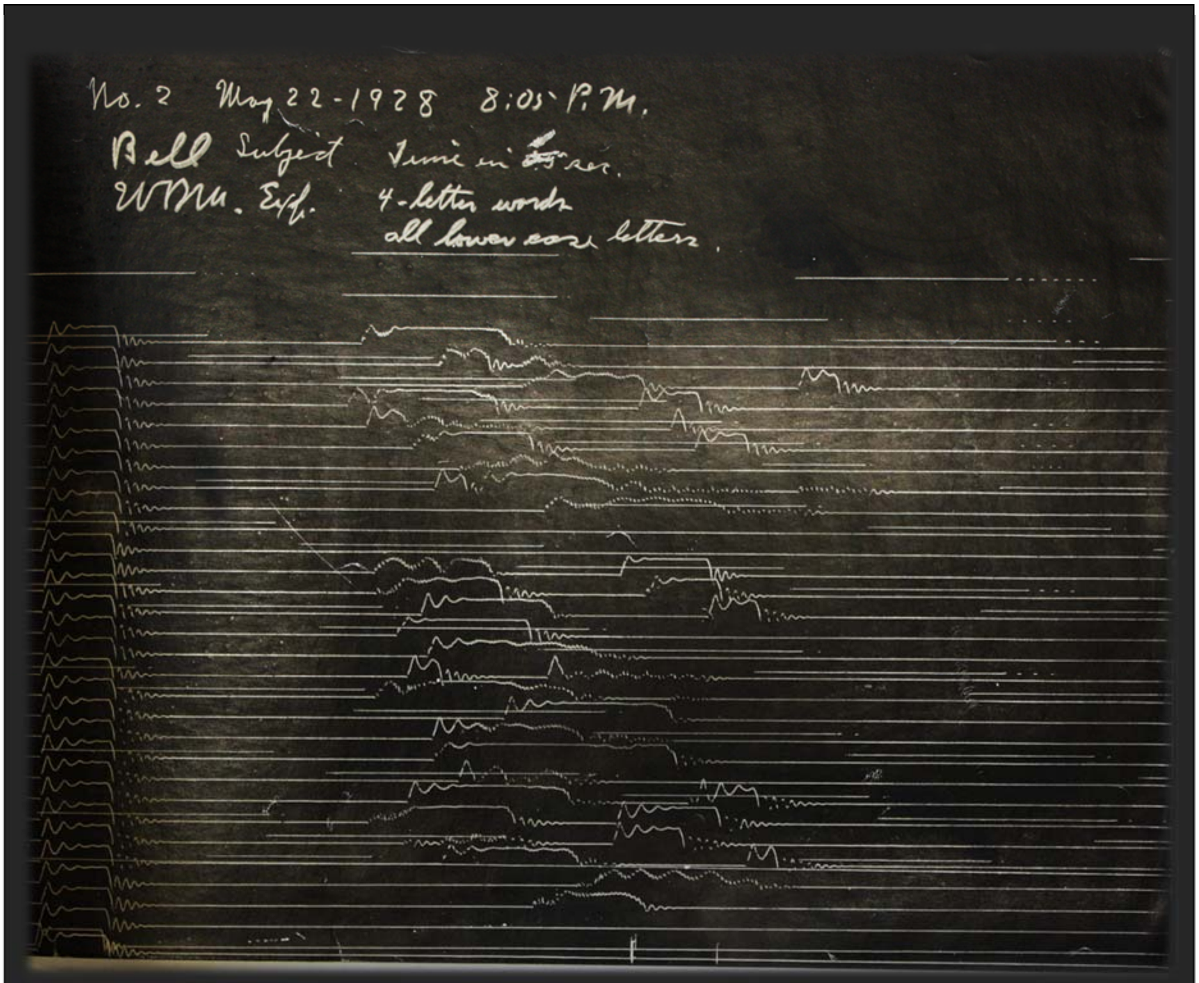

Figure 3.1: Example of kymograph paper with needle impressions. Each movement of the needle would be measured and converted to seconds. Unfortunately, we are unable to tell what the intertrial interval may have been for each trial.

of space. Alas, this is far below an optimal number for most techniques to detect the presence of 1/*f* in a timeseries. Rather than attempt to fit the data with any kind of frequency-based approach, we will see if the Wald-Wolfowitz runs test detects any ‘streakiness’ in the results. Testing each of the experimental sessions, each has fewer runs than would be predicted by chance; however, only one of these sessions has significantly less runs than expected and one is approaching significance. This is by

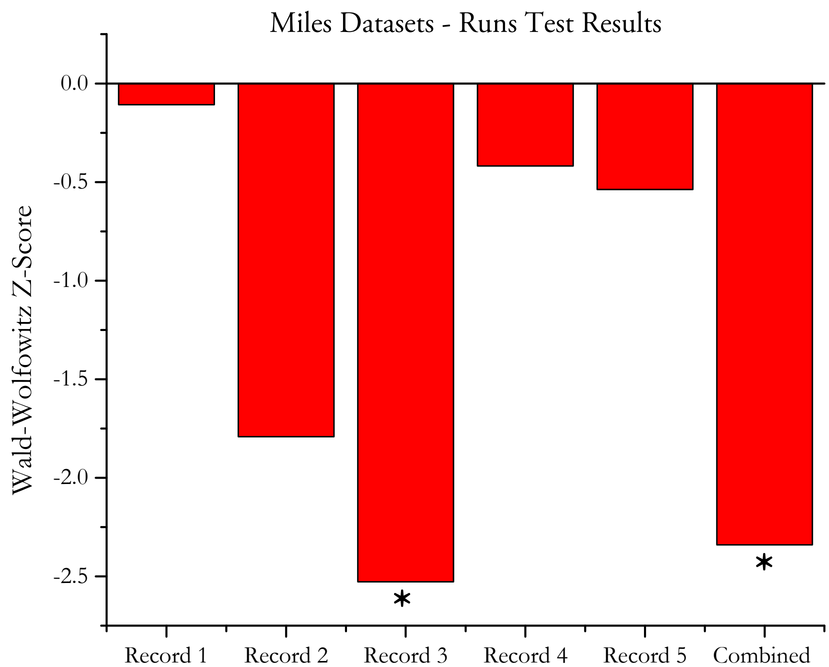

Figure 3.2: While looking at the records in the Miles Datasets, only one has significantly fewer runs than chance. When all records are concatenated into a single record, the resulting grand-record shows significantly fewer runs than expected.

no means a definitive result given the low number of samples in each session, but it does lean towards the hypothesis that previous generations also exhibited nonrandom (and possibly 1/*f*) patterns in their reaction times.

J. A. Morris Kimber

This dataset is from another student project that took place on July 16th, 1917 at the University of Pennsylvania. It is not known who taught this class, but the author can be traced from his start in undergraduate classes eventually earning a higher degree in clinical psychology. During WWII, Kimber enrolled in a graduate program

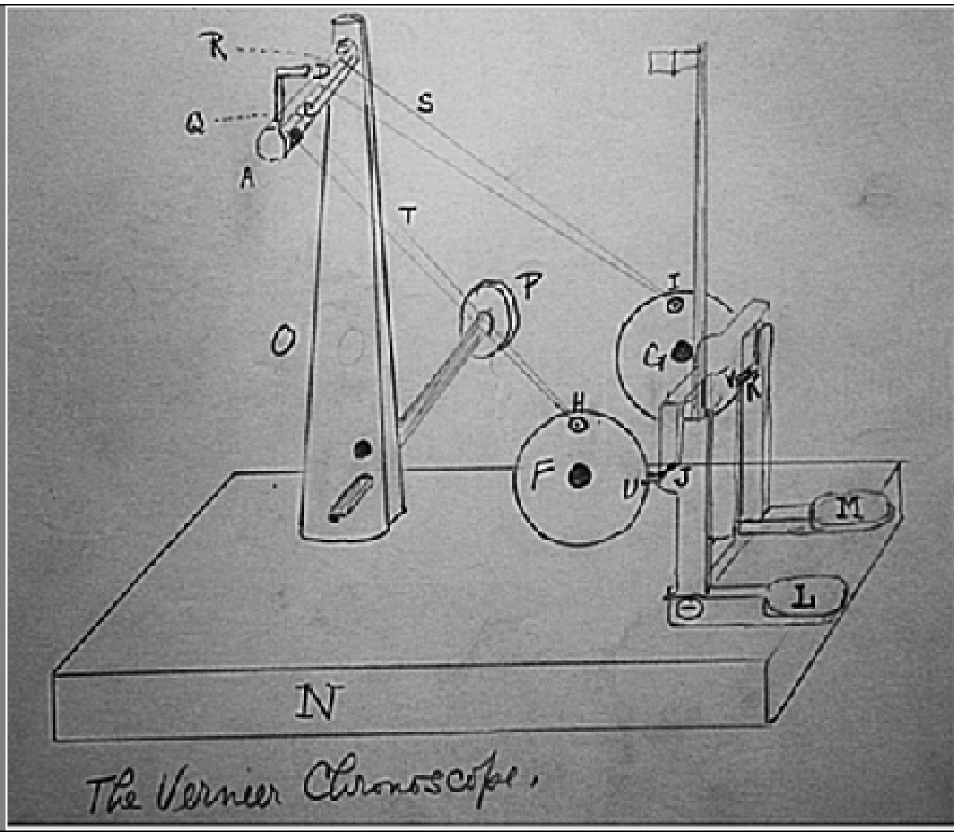

Figure 3.3: Kimber drew this schematic of the Vernier Chronoscope used for his class assignment. Note how the holes in the pendulums will allow a viewer to see a short flash when they momentarily line up.

at USC and earned a doctorate in abnormal and clinical psychology in 1945. Kimber's paper is from his time at Penn, and uses a Vernier chronoscope in order to investigate the difference between a simple reaction time to any stimulus as well as reaction time when having to listen for a specific stimulus.

At first glance, the Vernier chronoscope does not look like a precision measuring device, yet it can be a surprisingly accurate way to measure reaction time. It is comprised of two pendulums that are hung at different heights, and thus swing at different speeds. In Kimber's paper, one pendulum would swing back and forth every .78 seconds while the other would swing at .80 seconds. An experimenter would release the slower pendulum while displaying the stimulus and then the subject would respond by pressing a button that would release the faster pendulum. By counting the number of swings it took for the fast pendulum to catch up and align with the slow pendulum, the experimenter could calculate the reaction at roughly $1/50^{th}$ of a second accuracy. To put that rate in modern perspective, most USB keyboards are only accurate to $1/125^{th}$ of a second under ideal wired conditions, and can be much slower when used wirelessly.

In his project, Kimber records three sets of fifty reaction time trials. The first set is done simply as a reaction to the sound of the first pendulum being released. The second was done by listening between competing sounds and stopping between each trial to stop and write down introspections for each trial. The third set of trials was similar to the second, but the introspection was meant to be focused on the physical preparation of the finger striking the button. In accordance with David Gilden's work on which types of tasks show $1/f$ patterns (2001), we can hypothesize that recording simple reaction times like those in the first set of trials should be random while those involving more 'cognitive' mechanisms may be more $1/f$ and should show up as being more streaky using a Wald-Wolfowitz runs test.

The results are the complete opposite of the hypothesis. The first set of trials has significantly more runs than predicted while the last two trials do not even approach significance. Using a Whittle estimator, we can compare between combined sets with different sample sizes. The first set can be modeled as being slightly past $1/f$ (.64). The second and third sets are individually normed, concatenated, and modeled as being very close to random noise (.04). As noisy at these data are, it is not possible to conclude whether a power law itself is a good fit to the data.

It is difficult to come up with an interpretation of what these data mean. The pattern of runs and Whittle modeling being completely opposite of the hypothesis may either mean that our methodology is weak enough that the samples can go in an opposite direction, that we are receiving accurate information and the hypothesis is wrong, or perhaps the data were collected in such a way that would influence the outcome of the experiment. One way this might be is that, if the swinging and reset of the pendulum were to take several seconds, we may not be tapping into the actual endogenous rhythms of attention, but rather the changes happening due to the resetting of the chronoscope, which may take different amounts of time due to both the number of swings and efficiency resetting it. Regardless, this dataset raises more questions than it answers.

Wilson McTeer

Sometime between 1928 and 1930, a graduate student named Wilson McTeer at the University of Chicago would conduct a graduate-level class experiment that showed his competence using two different reaction time devices, the Vernier chronoscope (like that used by Kimber) as well as a newer and more expensive device, the Dunlap chronoscope. Knight Dunlap originally dubbed his creation the Johns Hopkins chronoscope when describing it in the literature (1917), but financial, philosophical, and moral conflicts involving Dunlap's former collaborator, John B. Watson, caused a rift that distanced him from the university. The chronoscope worked by utilizing the 60hz electrical frequency from a wall plug that could drive a motor

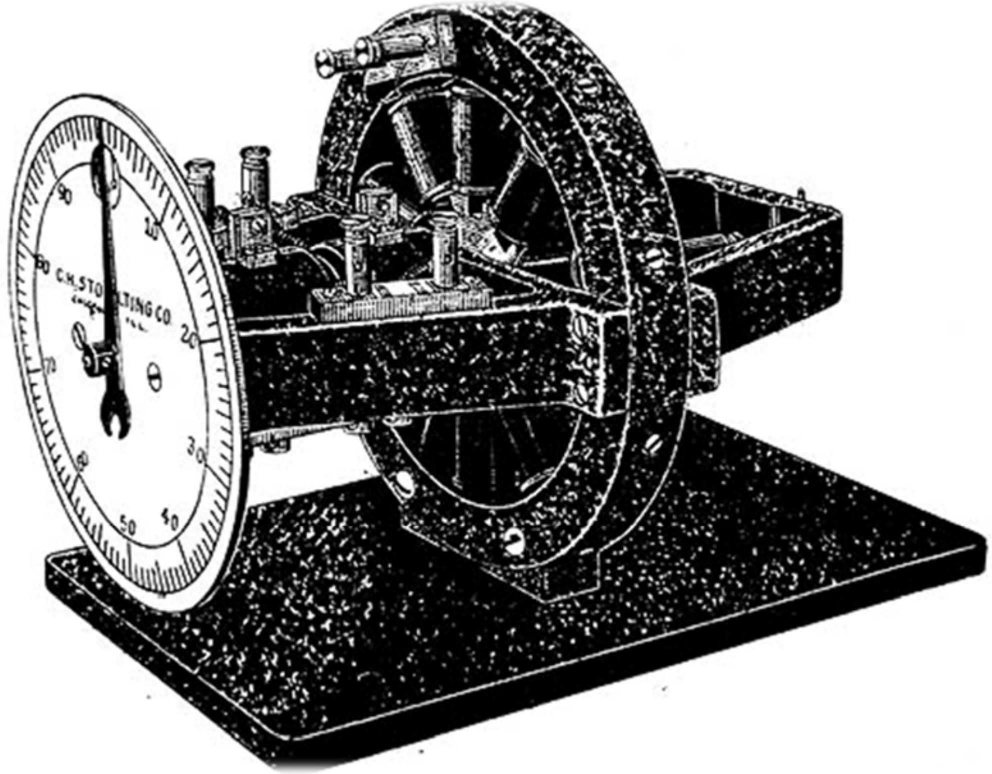

Figure 3.4: An image from an advertisement for the type of chronoscope developed by Dunlap and sold to psychologists (Stoelting, 1930, p. 83)

forward like an analog clock. Companies like C. H. Stoelting started producing the device for laboratories; however, these devices were rare because they were quite expensive, especially given the context of the 1929 stock market crash. An example of the Dunlap chronoscope can be seen in Figure 3.4. Wilson McTeer earned his PhD in 1930, and was hired immediately after commencement as a faculty member at the City College of Detroit. Throughout McTeer's career, he helped found the psychology department and help the college transition into becoming Wayne State University.

The exact methods for McTeer's assignment have been lost and are not part of the archive; however, a datasheet containing reaction times for multiple experiments and calculations can give us the raw reaction time numbers. The first data set comes from simple reaction times using the Dunlap chronoscope and sets of trials. These trials do not show any evidence of 'streakiness' using the Wald-Wolfowitz runs test. The other trials, taken with a Vernier chronoscope, are unfit for this type of analysis. Unfortunately, these are short datasets with only 25 trials and are clearly at floor, with subjects responding in, what appears to be, 100 to 140 milliseconds.

Consequently, this data does not shed a lot of light on whether previous patterns of attention were 1/*f* in nature.

Edmund Burke Delabarre

If there was ever an individual whose education prepared them to be an elite faculty member, it was Delabarre. Before founding the psychology laboratory at Brown University, he studied with William James (the father of American psychology) at Harvard, Hugo Muensterberg (the father of Industrial and Organizational Psychology) in Freiburg and Harvard, and Alfred Binet (the father of intelligence testing) at the Sorbonne. Despite appearing as a slam-dunk candidate for a job search committee, Delabarre was pretty much a bust, working for years on a book summarizing his work that was never completed. Instead, Delabarre became intensely interested in the Dighton Rock, a large rock with uninterpretable writing carved into it, which was conveniently located very close to his summer home. Delabarre believed the rock carving looked vaguely Portugese, a fact that pleased the government in Portugal. As a reward, he was named to the order of the St. James of the Sword, one of their highest honors for a non-national.

The data of interest for this dissertation was collected by Delabarre in 1902 for his a project on the influence of cannabis use. At the time, cannabis was not only legal but could be contained within medicines and sometimes was contained in medication completely unlabeled. Delabarre heroically took it upon himself to investigate what the psycho-motor effects of prolonged cannabis use might be, using himself as the one and only subject. On sporadic days (and occasionally twice a day) from Jun 19th to September 1st 1902, Delabarre would consume cannabis, run roughly sixty reaction time trials in four categories, and record the results. These data were never published, but Delabarre did manage to write roughly 300 words about his experimentation in an APA conference proceedings paper where he claimed that his ability to introspect was increased drastically when using the drug, and that it may have “enormous value in analysis.” (1899, p. 154). Delabarre does not say how he collected his reaction times; however, it may be possible that as a high-ranking professor he had access to a Hipp chronoscope, a device that uses precision clockwork driven by a dropping weight to measure time.

The Delabarre dataset does not provide enough trials to analyze in a single-sitting, but it does provide a series of times that are provided over a long timescale. By concatenating the individual datasets into a single, long dataset, we hopefully make comparisons between the two. The raw data was transcribed each day for two of the major categories - a simple reaction time task as well as a task that required remembering the association between two words. After transcription, there were 416 total trials for the simple task and 519 trials for the association task. Given the different numbers of trials, we will model the reaction times using the Whittle estimator. Surprisingly enough, the average reaction time for the more complex task is

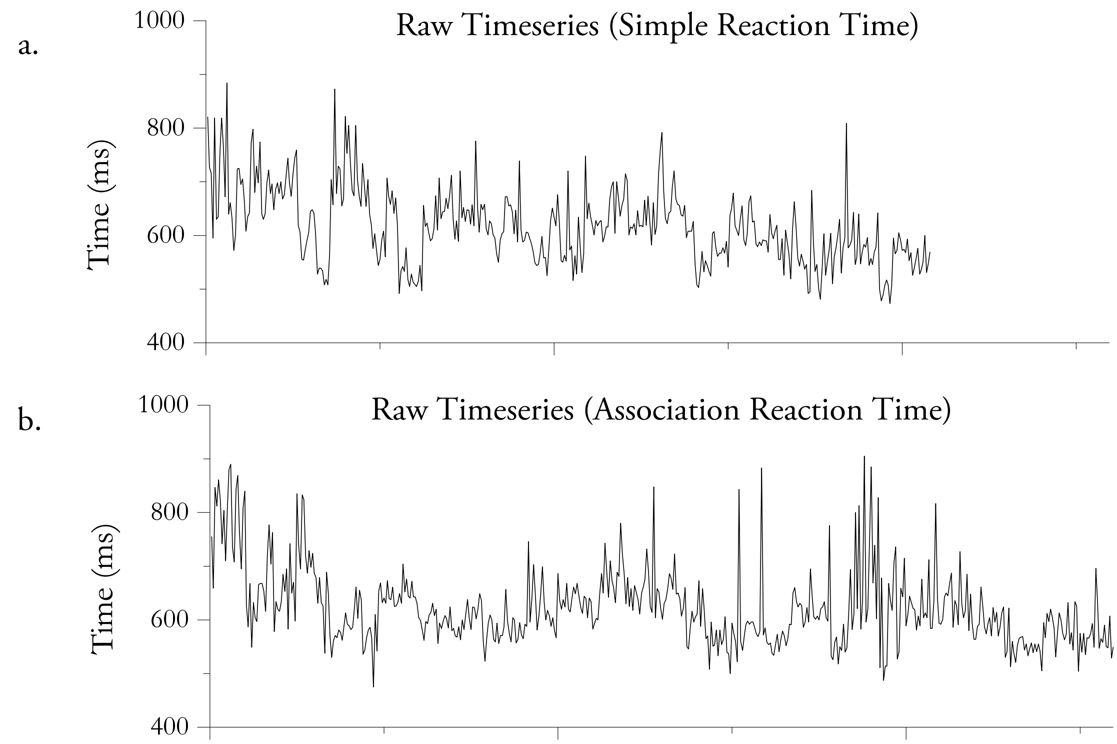

Figure 3.5: Graphs showing the concatenated series of raw reaction time data collected by Delabarre.

only an average of 6 milliseconds slower and not more statistically significant than the average for the simple task.

After modeling each timeseries using the Whittle estimator, the results do not fit as expected. The timeseries for the simple reaction time is modeled as having a higher dimensionality than $1/f$ ($\alpha = 1.22$) and the more complex task is slightly more than $1/f$ ($\alpha = 1.35$). As remarked in the previous chapter, successfully categorizing the difference between two timeseries with the observed difference in dimensionality throughout 500 trials is only about 70%. Therefore, this result, while the most statistically reliable in these historical datasets, is far from solid.

Even if this dataset has more statistical pull, other methodological details call into question whether this Delabarre's procedure is representative. The first and most obvious issue is that Delabarre was under the influence of cannabis while taking and administering the experiment. While his age and ongoing relationship with marijuana may explain why the reaction times are uncharacteristically long, it also means that his ability to adequately perform the experiment and record the results may have been compromised. In addition, Delabarre does not describe the specific method used to collect the data, so it may have been entirely inaccurate or inappropriate. Finally, the fact that the data are concatenated across multiple sessions may mean that any result observed is being driven more by day-to-day and strength-of-drug changes rather than anything we can pin down on attention.

Conclusions

Unfortunately, the data for this section of the dissertation are much like the joke of criticizing a bad meal: less than appetizing and with very small portions. Perhaps the most important takeaway of this project should be that data collected at the beginning of the 20th century probably would not be able to address the question of whether or not previous generations' patterns of attention were like ours. In the future, collecting data from the more recent past may prove to be more fruitful than looking for sources of reaction times that were collected before the advent of film, even if they have the confound of being influenced by changes in film.

What can be salvaged from the data? The Delabarre data appears to be the most robust, even if it was collected by a daily cannabis smoking, burntout professor. At the lowest level, we can detect that $1/f$ patterns did appear to exist in the reaction times at the beginning of the 20th century. The surprising part of the results are that those patterns cannot be consistently detected in the situations we would have predicted them to take place. Perhaps it will be necessary understand the function and structure of attention more completely in our own time before trying to make predictions about what the past may have looked like.

CHAPTER 4

Visual Memory and Attention for Self-Paced and Induced 1/*f* Timeseries

Introduction

In retrospect, identifying the presence of 1/*f* noise found in reaction times (Gilden, 1995, 1997, 2001) might not be a particularly surprising discovery. The 1/*f* pattern has been investigated and modeled in most corners of our world including in the structure of natural scenes (Field, 1987), traffic jams (Nagatani, 1993), the career trajectory of hotel executives (Houran, Lange, & Kefgen, 2013), and biological systems in general (Musha & Yamamoto, 1997; Szendro, Vincze, & Szasz, 2001). The question remains, why is this type of noise ubiquitous? Is it indicative of some sort of meaningful and functional underlying structure or is it simply a consequence of some other factor? Some have claimed that 1/*f* noise in the brain is an epiphenomena of the way our brains are physically structured, and doesn't serve any sort of functional role in cognition (Christopher M Anderson et al., 1993). This point of view explains the variation found in a variety of tasks from movement to lexical processing (Van Orden, Holden, & Turvey, 2003). Others have posited that 1/*f* noise is more than a byproduct, playing an active role in cognitive processes such as the emergence of consciousness (Carl M Anderson & Mandell, 1996) and the linking of cognition and emotion (Carl M Anderson, 2000).

Given our finding that the editing patterns in Hollywood film are slowly evolving towards a 1/*f* pattern (Cutting et al., 2010) we proposed that the change could represent that filmmakers are unknowingly taking advantage of the endogenous rhythms found in human attention. Does the 1/*f* structure found in attentional and cognitive processes actually influence performance, or is it simply a byproduct of biology? Can a temporal pattern that is observed endogenously have an influence on memory and attention when it is presented as part of an exogenous stimulus? The following experiments will attempt to look for behavioral effects of 1/*f* stimuli.

Experiment 1: Self-Paced Visual Presentation

Self-paced presentation is rarely the norm in psychology experiments. Most experiments fixed interval stimuli presentations with the occasional random inter-trial interval in order to keep subjects from being able to completely anticipate when a trial will begin and end. The reasons for using fixed intervals are very straightforward, as replication is easier when subjects are shown the exact same stimulus for the exact same interval. It also just feels more scientific and precise to hold everyone to the same script when evaluating performance. The downside to using this approach is that a fixed interval of presentation has little to no bearing with anything found in the natural world. With the possible exception of decaying elements and spinning stars, the world ebbs and flows with changes during decades, seasons, and minutes. People are no exception, and research has shown that events in contemporary film and events in life (even those events are as artificial as a basketball game) share similar temporal structure (Blau, Petrusz, & Carello, 2013). This fundamental disconnect between the temporal structure of the world and our experiments have lead some researchers to call into question our de-facto methodology, as anticipation in fixed-interval patterns differs from anticipation with biological rhythms (Torre, Varlet, & Marmelat, 2013).

The idea that people may be able to synchronize with the timeseries properties of a Hollywood film might not be as farfetched as it sounds. Studies have observed that when attempting to tap along with different types of chaotic timeseries, people are able to synchronize their taps with long-range correlation changes in the signal (Stephen, Stepp, Dixon, & Turvey, 2008). While we might not be completely aware of the temporal structure of events around us, we do appear to be sensitive to them. This experiment will attempt to see what type of structure is apparent in subjects performing a self-paced task, and whether that structure has an influence on a subject's memory.

Experiment 1: Methods

Two groups of undergraduate students participated in the experiment. The first group of seventeen subjects was recruited online and participated in the experiment for course credit. The second group of nineteen subjects was recruited online and participated in the experiment in exchange for payment. All subjects were informed of the parameters of the experiment and signed papers acknowledging that their participation was voluntary.

Subjects were instructed on-screen that they would be participating in an experiment on picture memory. They were instructed to go through a randomly generated slideshow of 260 images and attempt to remember as many as possible for a later of the experiment. The sequence of images was a randomly selected and ordered subset of greyscale images taken from the Van Hateran database (van Hateren & van der Schaaf, 1998), that were converted to 8-bit images, cropped randomly to be square, and rescaled to be a 256x256 pixel image. The images were collected as

Figure 4.1: Representative images from the van Hateran database (1998) after cropping and converting images.

examples of natural scenes, and are mostly comprised of foliage, flowers, and shrubbery as well as some infrequent buildings, people, and signage. Representative images from the dataset can be seen in Figure 4.1. The amount of time that subjects viewed each image as well as their sequence and image number were stored.

After self-paced viewing of the slideshow, subjects received on-screen instructions that they would now be asked to recognize which of two images they had previously seen on the screen. A novel image was paired with a previously viewed image and presented side-by-side for one second, with side being assigned randomly. Once the images had left the screen, subjects used a keypress (either 'F' or 'J') to indicate which image they had seen before, and responses were recorded for 100 trials.

The primary goal of this study is to determine whether temporal structure of the study phase has any effect on accuracy and recall. This requires the assumption that increased attention in the study phase should increase the subject's ability to recall each picture in memory. In order to model the timeseries during the study phase, we can use a Whittle estimator to check if the subject displayed a temporal pattern that is closer to white, pink, or brown noise.

Experiment 1: Results

A binomial distribution (n=100, p = .5) was used to determine what scores could be expected by chance with 99% confidence, and subjects with a score lower than 63% were excluded from the results. This excluded three subjects from the first sample and twelve subjects from the second study. The discrepancy between the groups may be attributable to the fact that the subjects being paid to participate in this long, boring study were less likely to maintain vigilance than the students who volunteered for class. Due to the low number of subjects performing better than chance in the second sample, the samples will be merged and analyzed together in addition to individually analyzed.

The first four study trials for each subject were trimmed to allow subjects a few trials to acclimate to the study, leaving 256 entries available to model with a Whittle estimator. Timeseries were transformed into z-scores and detrended before analysis. Results from the two individual sets of subjects and from the combined dataset shows a positive correlation between the dimensionality (in α) of the study-phase timeseries and performance on the memory task ($R^2 = .189$, see Figure 4.2). Poorly performing subjects were modeled as having a range of dimensionality values; however, those who performed well at the task were all modeled as being different from a random timeseries.

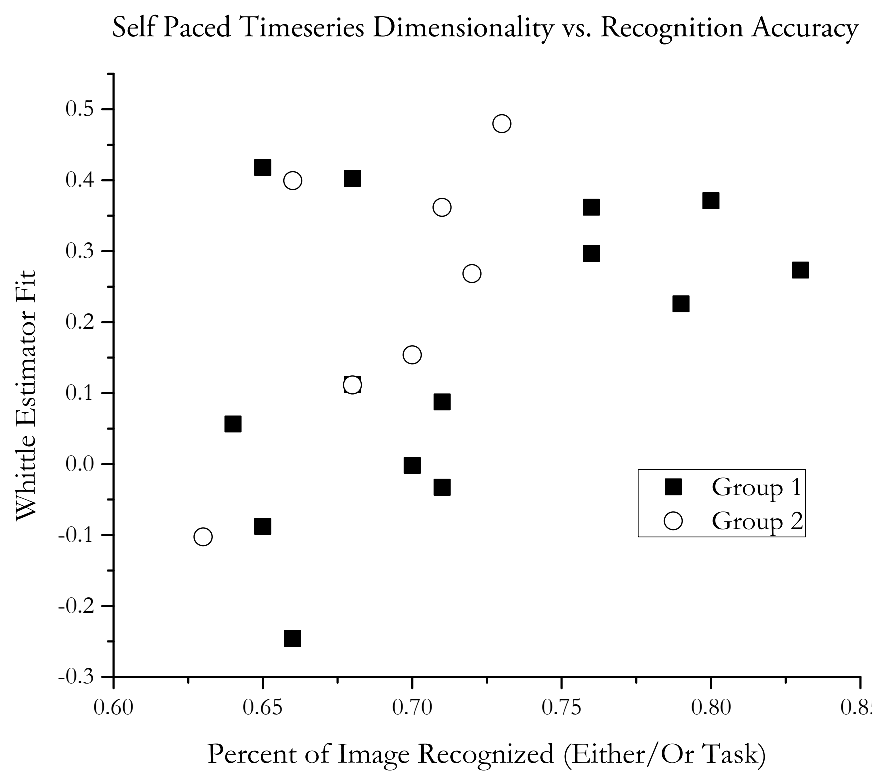

Figure 4.2: Results from two experimental groups show a positive correlation between dimensionality and percent of images recognized.

In addition to the dimensionality (in α) of the timeseries, an important aspect of the study-phase to analyze is the amount of time that subjects spent looking at each image. It would make sense that the subjects who spent more time studying each image should perform better in the recognition part of the experiment. Figure 4.3 shows it is true that greater self-presentation time correlates with recognition accuracy, though this effect is not significantly different than chance ($R^2 = .062$)

While the results gleaned from these data may seem cut and dried, the fact is that if both the mean viewing time and dimensionality fit are entered into a linear regression, neither comes out as significant. In addition, if values below the cutoff value of 63% are included in the sample the dimensionality fit is no longer a significant predictor of accuracy. While it makes sense that subjects who are not paying attention or putting effort into the task will not show effects, the fact that, many of them showed a different-than-random outcome means that $1/f$ clearly is not restricted to when subjects are paying attention to a stimulus.

Experiment 1: Discussion

While this study shows promise towards linking a $1/f$ pattern of study and enhanced recognition memory in a subsequent task, it has a number of potential improvements before it is ready to be published. The first issue is that fifteen of thirty-six subjects were excluded from the study, which is a rate much larger than predicted when first planning the experiment. Subsequent studies have attempted to incentivize subjects by rewarding correct answers rather than pay at a flat rate. At the same time, a slideshow of 260 greyscale images may be just too boring for subjects to take in, which is in direct opposition of the second issue with this study: that the timeseries may need to be longer for the Whittle estimator's fits to stand up to peer review.

With 256 data points in a timeseries, there simply may not be adequate data to model dimensionality with enough accuracy. Given the uniqueness of this approach, reviewers simply will not understand accuracy issues with this type of timeseries analysis. However, increasing the number of trials may be difficult as subjects completing the task above the accuracy threshold took roughly fourteen minutes to get through the 260 trials. Asking subjects to remain engaged with the task of memorizing pictures of trees and shrubs for more than fifteen continuous minutes may be too much to ask, though this seems surprising when some studies have recorded subjects recognizing slight changes in image state with better than 80% accuracy (Brady, Konkle, Alvarez, & Oliva, 2008). Another option is to include a break to allow subjects to recompose themselves, although a gap in the timeseries could open the modeling up to criticism.

It is possible to look at the results of this study and perform a power analysis (Faul, Erdfelder, Buchner, & Lang, 2009) which, in addition to describing the odds of finding the current set of results, can help hone in on what it would take for the next experiment to have a more-than-reasonable chance of succeeding. Converting both subject group's R^2 value into an F^2 value, it is possible to calculate the odds of finding a significant result were for the first study given the effect size observed and number of subjects. Modeling the results as a multiple regression with one predictor, the estimated post-hoc probability is at 55%, which is not unreasonably underpowered, but can be improved. In order to achieve a power level where the study is capable of returning a significant finding 90% of the time, the number of subjects needs to be more than doubled, with the data from 48 subjects entered into the regression. Given that 42% of subjects did not perform above the chance cutoff, this future study should recruit roughly 70 subjects.

Experiment 2: Recall Memory for Fixed vs. 1/f Temporally Displayed Images

While the first study in this chapter helps to suggest that a $1/f$ pattern in self-presentation times may have something to do with subsequent memory, this does not help support our conjecture that the $1/f$ patterns in film have functional significance to the audience. Instead, it may be the case that $1/f$ patterns in film are simply an artifact of filmmakers efforts and do not influence the audience at all. In order to test whether exogenous temporal patterns influence memory, we can repeat the previous experiment using pre-set temporal patterns. By showing subjects similar images in an either fixed-rate or 1/*f* slideshow, we can determine if the temporal pattern is influencing memory.

Experiment 2: Methods

Fifty-seven subjects were recruited in three stages for this experiment. All subjects were undergraduates participating for course credit. The protocol, images, and details of experiment were identical to the previous experiment with the major exception that subjects could not control when the images were presented. In the fixed condition, a randomized selection of Van Hateran images were presented for two seconds. In the 1/*f* condition, presentation time for each image was determined by following a random 1/*f* sequence, generated in MATLAB (2007) with the average presentation time at two seconds. Each random sequence was modeled using Gilden's method before presentation in order to determine that the sequence was representative of 1/*f* noise (alpha between .95 and 1.05).

Experiment 2: Results

Overall, subject recognition memory was slightly but not significantly higher ($p = .15$) for the 1/*f* sequence in comparison to the fixed sequence. Subjects did not perform particularly well at the task especially confronted with the 63% threshold established in the first experiment; only 40% of the subjects were above threshold in the fixed presentation condition and only 44% were above the threshold for the 1/*f* presentation condition.

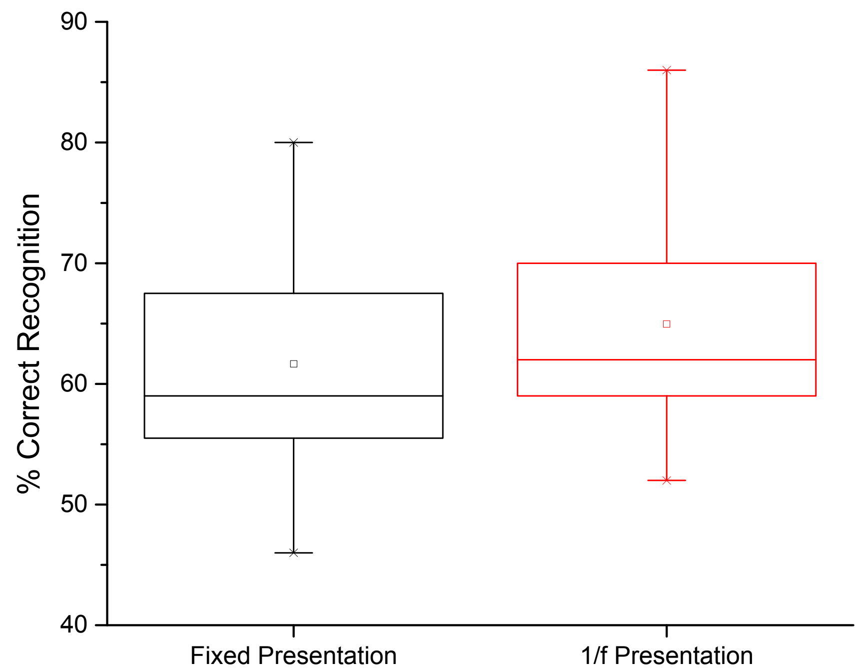

Figure 4.3: Although subjects were better at recognizing images in the 1/*f* presentation versus the fixed presentation, the difference was not statistically significant.

Experiment 2: Discussion

While the results of this study are somewhat underwhelming, the data do give reason to pursue a more complete study. While there is no significant difference in the two groups, there is a trend in the predicted direction. As seen in Figure 4.3, the difference in groups appears as a shift of a slightly skewed distribution rather than due to outliers or different variances between groups. Although this experiment did not achieve statistical significance, the Cohen's d calculated for the study was .39, a medium-sized effect. By utilizing the effect and sample sizes, we can see that if the effect exists in the size predicted, the current experimental setup would only show statistical significance 29.9% of the time. Replicating the experiment to achieve statistical significance 90% of the time would require a much larger sample, with 141 subjects in each condition. This amount can be cut in half by using a within-subjects design, however that also doubles the length of the experiment, which is already half an hour in length.

Conclusions

An early and major milestone in psychology was Hermann Ebbinghaus' work on distributed and massed practice (Ebbinghaus, Ruger, & Bussenius, 1913). Through his studies, Ebbinghaus would test himself by writing down a series of pseudo-words and then attempting to recite them to the beat of a metronome. Not only did this research succeed in Ebbinghaus' goal to show that higher mental processes could be studied experimentally, it also gave us two concepts that continue to be relevant today: the forgetting curve and the spacing effect. These concepts are cornerstones in the study of attention and memory, yet we have not truly investigated why they occur. It is taken for granted that we tend to forget things at a roughly exponential decay, but why is this the case? The way information is interfered-with and forgotten over time *tells us something meaningful* about either the structure of our brains or the structure of the world our brains live in.

The concepts of cramming and spacing are well known in psychology, and are typically invoked in order to sing the praises of spacing and shame our students who engage in cramming. Despite Ebbinghaus' results, hundreds of researchers have felt the need to replicate the spacing effect in different domains including areas such as training typists (Baddeley & Longman, 1978). Only recently have researchers tried to take a look inside the brain and figure out why this ubiquitous truth of spacing appears, over and over again, in our literature (Kramár et al., 2012). Despite all the literature, it is important to note that spacing and cramming were never meant to be the only two temporal structures of learning worth studying. They were a convenient way for Ebbinghaus to test himself, but they do not represent the only two options for learning over time.

Although the experimental studies contained in this chapter are not likely to set the world on fire, they represent a first step in a line of research that I expect to have some impact on the status quo of psychology. Attention and memory are typically studied a couple hundred milliseconds at a time, and I blame William James for influencing psychologists to believe that anything above a couple seconds isn't worth studying. We have seen glimpses of a world past a couple seconds in the studies of Mackworth, LaVie, and Gilden, but these studies have not pushed us beyond thinking of attention as short bursts of intense, sustained effort directed willfully towards the world.

While it can be useful to study attention in laboratories, straining our resources until we find our breaking point, the truth is that in the real world attention does not take the form of memorizing an array of numbers and letters. Attention is a process that is constantly being used perceptually to guide our eyes to probabilistically relevant features of the environment. As we listen to words around us, our eyes are making bets about what may be the most relevant information in near real-time (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This type of attention, which occurs nearly every second of every day, occurs without sustained effort, and in many ways looks like a different process than most laboratories study. I would argue that it is this type of attention (the one that stitches together and makes sense of the world) is active when we are watching movies.

What do we have to gain by studying a less impressive, more pedestrian form of attention? I would argue that the act of actively, effortfully, and consciously focusing attention is actually a relatively strange and abnormal state to put attention into. This form of attention is brittle and short-lasting, and realistically only needs to be rolled out during early learning or in a high arousal situation. When we test sustained bursts of effortful attention, we are measuring what the system *can* do, but not necessarily what it *does*. To force analogies on the situation, taking a system that typically works in the background and make it effortfully engage in focusing for long stretches is akin to judging the capability of a long distance runner by making them run a 40-yard dash or deciding car of the year by off-roading in a Prius. Simply because your attention *can* effortfully focus to a location outside your foveal vision does not mean that is what the system actually develops to do.

It is my hope that my future work will show evidence that there is more structure to our attention than we give it credit for. As stated, $1/f$ structure and lognormal distributions in behavioral data may be so ubiquitous that they are not particularly special. Even if this is the case, it is still worth talking about what type of system would produce these patterns with such surprising regularity. I think that in order to ask questions about the multilevel structure of the mind, we need to start moving away from previous conceptions of the brain. I believe researchers lean too heavily on the computer metaphor and information theory. We may be better served by using new conceptions of the brain as complex system that does not have the goal of creating neatly organized systems we can draw lines around, but rather individual parts that when put together form basic organizing principles, such as receptive fields, distributed representations, sparsity, and Hebbian connectionism, which can be used over and over again in the brain. Isolating how some of the population level structure of the brain is being changed we may be able to get a better idea of what principles are being used and recruited in system organization.

REFERENCES

Aitchison, J., & Brown, J. A. (1957). *The Lognormal Distribution* (1st ed.). Cambridge, UK: Cambridge University Press.

Akaike, H. (1980). Likelihood and the Bayes procedure. *Trabajos de Estadística Y de Investigación Operativa*, *31*(1), 143–166.

Anderson, C. M. (2000). From molecules to mindfulness: How vertically convergent fractal time fluctuations unify cognition and emotion*. *Consciousness & Emotion*, *1*(2), 193–226.

Anderson, C. M., Holroyd, T., Bressler, S. L., Selz, K. A., Mandell, A. J., & Nakamura, R. (1993). 1/f-like spectra in cortical and subcortical brain structures: A possible marker of behavioral state-dependent self-organization. In *Noise in physical systems and 1/f fluctuations* (Vol. 285, pp. 737–740). AIP Publishing.

Anderson, C. M., & Mandell, A. J. (1996). Fractal time and the foundations of consciousness: Vertical convergence of 1/f phenomena from ion channels to behavioral states. *Fractals of Brain, Fractals of Mind*, *7*, 75–126.

Anderson, J. D. (1998). *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Southern Illinois University Press.

Ariga, A., & Lleras, A. (2011). Brief and rare mental "breaks" keep you focused: deactivation and reactivation of task goals preempt vigilance decrements. *Cognition*, *118*(3), 439–43. doi:10.1016/j.cognition.2010.12.007

Aserinsky, E., & Kleitman, N. (1953). Regularly Occurring Periods of Eye Motility, and Concomitant Phenomena, During Sleep. *Science*, *118*(3062), 273–274.

Bacevich, A. J. (2013, March). A Letter to Paul Wolfowitz. *Harper's Magazine*, 48–50.

Bachmann, B., & Huston, J. (1965). How I make films: An interview with John Huston. *Film Quarterly*, *19*(1), 3–13.

Baddeley, A. D., & Longman, D. J. A. (1978). The Influence of Length and Frequency of Training Session on the Rate of Learning to Type. *Ergonomics*, *21*(8), 627–635. doi:10.1080/00140137808931764

Barber, N. (2006). Is the effect of national wealth on academic achievement mediated by mass media and computers? *Cross-Cultural Research*, *40*(2), 130–151.

Basketball Reference. (2014). Darryl Dawkins. *BasketballReference.com*. Retrieved February 06, 2014, from http://www.basketball-reference.com/players/d/dawkida01.html

Baxter, M. (2013). Evolution in Hollywood editing patterns? *Cinemetrics Website*. Retrieved June 22, 2014, from http://www.cinemetrics.lv/dev/Evolution_paper_for_Cinemetrics.pdf

Blau, J. J. C., Petrusz, S. C., & Carello, C. (2013). Fractal Structure of Event Segmentation: Lessons From Reel and Real Events. *Ecological Psychology*, *25*(1), 81–101.

Bordwell, D. (2006). *The Way Hollywood Tells it: Story and Style in Modern Movies*. University of California Press.

Bossom, J., Natelson, B. H., Levin, B. E., & Stokes, P. E. (1983). Ultradian rhythms in cognitive functions and their relationship to visceral processes. *Physiology & Behavior*, *31*(1), 119–123. doi:10.1016/0031-9384(83)90105-1

Bottomore, S. (1990). Shots in the dark: the real history of film editing. In T. Elsaesser & A. Barker (Eds.), *Early Cinema: space frame narrative* (pp. 104–113). London: BFI Publishing.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

Broadbent, D. E. (1958). *Perception and communication* (p. 338). Oxford: Pergamon Press.

Bryce, R. M., & Sprague, K. B. (2012). Revisiting detrended fluctuation analysis. *Scientific Reports*, *2*.

Cahalane, D. (2013, August 19). *Mathematical Models For The Evolution And Development Of The Cerebral Cortex In Mammals*. Retrieved from http://ecommons.library.cornell.edu/handle/1813/34161

Carroll, N. (1998). The Future of Allusion: Hollywood in the Seventies (and Beyond). In *Interpreting the Moving Image* (pp. 51–81). Cambridge: Cambridge University Press.

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975. doi:10.1121/1.1907229

Conte, S., Ferlazzo, F., & Renzi, P. (1995). Ultradian rhythms of reaction times in performance in vigilance tasks. *Biological Psychology*, *39*(2-3), 159–72. doi:10.1016/0301-0511(94)00962-W

Correll, J. (2008). 1/f noise and effort on implicit measures of bias. *Journal of Personality and Social Psychology*, *94*(1), 48.

Correll, J. (2011). Order from chaos? 1/f noise predicts performance on reaction time measures. *Journal of Experimental Social Psychology*, *47*(4), 830–835. doi:http://dx.doi.org/10.1016/j.jesp.2011.02.019

Coullet, P., & Tresser, C. (1978). Iterations d'endomorphismes et groupe de renormalisation. *J. Phys. Colloques*, *39*(C5), C5–25–C5–28. Retrieved from http://dx.doi.org/10.1051/jphyscol:1978513

Cutting, J. E. (2005). Perceiving Scenes in Film and in the World. In J. D. Anderson & B. F. Anderson (Eds.), *Moving image theory: Ecological considerations* (pp. 9–27). Carbondale, IL: Southern Illinois University Press.

Cutting, J. E. (2014). More on the Evolution of Popular Film Editing. *Cinemetrics Website*. Retrieved from http://www.cinemetrics.lv/dev/cuttingcinemetricx3.pdf

Cutting, J. E., Brunick, K. L., & DeLong, J. E. (2011). The changing poetics of the dissolve in Hollywood film. *Empirical Studies of the Arts*, *29*(2), 149–169.

Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *I-Perception*, *2*(6), 569.

Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science*, *21*(3), 432–439. doi:10.1177/0956797610361679

Delabarre, E. B. (1899). Proceedings of the seventh annual meeting of the American Psychological Association, Columbia University, New York, December, 1898. *Psychological Review*, *6*(2), 146–179. Retrieved from 10.1037/h0071957

DeLong, J. E. (2013). Horseshoes, handgrenades, and model fitting: the lognormal distribution is a pretty good model for shot-length distribution of Hollywood films. *Literary and Linguistic Computing*.

DeLong, J. E., Brunick, K. L., & Cutting, J. E. (2012). Film through the Human Visual System: Finding Patterns and Limits. In *Social Science of the Cinema*. Oxford University Press.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80.

Donders, F. C. (1867). Deux instruments pour la mesure du temps nécessaire pour les actes psychiques. *Archives Néerlandaises Des Sciences Exactes et Naturelles*, *2*, 247–250.

Dunlap, K. (1917). The Johns Hopkins chronoscope. *Journal of Experimental Psychology*, *2*(3), 249.

Ebbinghaus, H. (1911). *Grundzüge der Psychologie.* Leipzig: Veit.

Ebbinghaus, H., Ruger, H. A., & Bussenius, C. E. (1913). *Memory: A Contribution to Experimental Psychology* (Educationa.). New York City: Teachers College, Columbia University.

Elam, J. (2011). Vice President Joe Biden Highlights the HCP at Event to Support Brain Research. *Human Connectome Project*. Retrieved May 21, 2014, from http://humanconnectome.org/about/pressroom/project-news/vice-president-joe-biden-highlights-hcp-event-support-brain-research/

Farrell, S. (2008). PBRmatlab code. *Professional Website*. Retrieved January 07, 2014, from http://eis.bris.ac.uk/~pssaf/PBRmatlab.zip

Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2006). 1/f noise in human cognition: Is it ubiquitous, and what does it mean? *Psychonomic Bulletin & Review*, *13*(4), 737–741. doi:10.3758/BF03193989

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fechner, G. T. (1860). *Elemente der Psychophysik. 1*. Leipzig: Breitkopf & Härtel.

Federal Aviation Administration. Doc. No. FAA-2009-1093. , Pub. L. No. 77 FR 28764 (2012).

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, *4*(12), 2379–2394.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*(1), 29.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171.

Gallos, L. K., Sigman, M., & Makse, H. A. (2012). The Conundrum of Functional Brain Networks: Small-World Efficiency or Fractal Modularity. *Frontiers in Physiology*, *3*, 123. doi:10.3389/fphys.2012.00123

Galton, F. (1883). *Inquiries Into Human Faculty and Its Development*. London: Macmillan.

Gibson, J. J. (1979). *The ecological approach to visual perception:*. Houghton Mifflin.

Gilden, D. L. (1995). 1/f Noise in Human Cognition. *Science*, *267*(5205), 1837–1839. doi:10.2307/2886349

Gilden, D. L. (1997). Fluctuations in the Time Required for Elementary Decisions. *Psychological Science*, *8*(4), 296–301. doi:10.2307/40063196

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, *108*(1), 33–56. doi:10.1037/0033-295X.108.1.33

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314.

Gopher, D., & Lavie, P. (1980). Short-term rhythms in the performance of a simple motor task. *Journal of Motor Behavior*, *12*(3), 207–219.

Gottschalk, A., Bauer, M. S., & Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry*, *52*(11), 947–959. doi:10.1001/archpsyc.1995.03950230061009

Graham, D. J., & Field, D. J. (2008). Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, *21*(1-2), 149–164.

Helton, W., & Russell, P. (2012). Brief mental breaks and content-free cues may not keep you focused. *Experimental Brain Research*, *219*(1), 37–46. doi:10.1007/s00221-012-3065-0

Hilgard, E. R. (1985). Walter Richard Miles: 1885-1978. In *National Academy of Sciences Memoirs* (pp. 409–432). Washington, D.C.: National Academy of Sciences.

Hobbs, R., Frost, R., Davis, A., & Stauffer, J. (1988). How First-Time Viewers Comprehend Editing Conventions. *Journal of Communication*, *38*(4), 50–60. doi:10.1111/j.1460-2466.1988.tb02069.x

Hochberg, J. (1970). Attention, organization, and consciousness. In D. Mostofsky (Ed.), *Attention: Contemporary Theory and Analysis*. Appleton-Century-Crofts.

Holland, J. H. (2006). Studying Complex Adaptive Systems. *Journal of Systems Science and Complexity*, *19*(1), 1–8. doi:10.1007/s11424-006-0001-z

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 1–23.

Houran, J., Lange, R., & Kefgen, K. (2013). Industry Trends: Fascinating Rhythms in the Career Paths of Hospitality Executives . *Cornell Hospitality Quarterly* , *54* (1), 6–9. doi:10.1177/1938965512467602

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, *116*, 770–808.

Hurst, H. E., Black, R. P., & Simaika, Y. M. (1965). *Long-term storage: an experimental study*. London: Constable.

Ildirar, S., & Schwan, S. (2014). First-time viewers' comprehension of films: Bridging shot transitions. *British Journal of Psychology*.

IMDb Inc. (2014). The Internet Movie Database (IMDb). Retrieved from http://www.imdb.com/

James, W. (1901). *The principles of psychology.* London: Macmillan.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs: Prentice-Hall Inc.

Kantor, J. R. (1922). Can the Psychophysical Experiment Reconcile Introspectionists and Objectivists? *The American Journal of Psychology*, *33*(4), 481–510. doi:10.2307/1413546

Kelly, F. P. (1994). *Probability, statistics and optimisation: A Tribute to Peter Whittle* (p. 5). Chinchester: John Wiley & Sons.

Klein, J. (1997). *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938* (p. 268). Cambridge, UK: Cambridge University Press.

Klein, R., & Armitage, R. (1979). Rhythms in Human Performance: 1½-Hour Oscillations in Cognitive Style. *Science*, *204*(4399), 1326–1328. doi:10.2307/1748768

Kleitman, N. (1939). *Sleep and Wakefulness as Alternating Phases in the Cycle of Existence*. University of Chicago Press.

Kleitman, N. (1954). The Sleep-Wakefulness Cycle in Submarine Personnel. In *A Survey Report on Human Factors in Undersea Warfare* (pp. 328–341). (U.S.), National Academy of Sciences, (U.S.), National Research Council.

Kleitman, N. (1963). *Sleep and wakefulness*. University of Chicago Press.

Kramár, E. A., Babayan, A. H., Gavin, C. F., Cox, C. D., Jafari, M., Gall, C. M., … Lynch, G. (2012). Synaptic evidence for the efficacy of spaced learning. *Proceedings of the National Academy of Sciences* , *109* (13), 5121–5126. Retrieved from http://www.pnas.org/content/109/13/5121.abstract

Kuensch, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the first World Congress of the Bernoulli Society* (Vol. 1, pp. 67–74). VNU Science Press Utrecht.

Lavie, P. (1979). Ultradian rhythms in alertness — A pupillometric study. *Biological Psychology*, *9*(1), 49–62. doi:10.1016/0301-0511(79)90022-X

Lavie, P. (1980a). The search for cycles in mental performance from Lombard to Kleitman. *Chronobiologia*.

Lavie, P. (1980b). Ultradian rhythmicity (circa 90 min) in physiologic processes-evidence for multioscillatory phenomena. *Biological Rhythm Research*, *11*(3), 171–177.

Lavie, P. (1992). Ultradian Cycles in Sleep Propensity: Or, Kleitman's BRAC Revisited. In D. Lloyd & E. Rossi (Eds.), *Ultradian Rhythms in Life Processes SE - 13* (pp. 283–302). Springer London. doi:10.1007/978-1-4471-1969-2_13

Lavie, P., & Zomer, J. (1984). Ultrashort sleep-waking schedule. II. Relationship between ultradian rhythms in sleepability and the REM-non-REM cycles and effects of the circadian phase. *Electroencephalography and Clinical Neurophysiology*, *57*(1), 35–42.

Levin, D., Ildirar, S., Schwan, S., & Smith, T. (2014). The role of audio on perception of continuity by first-time viewers. In *The Society for Cognitive Studies of the Moving Image*. Talk Presented at SCSMI 2014, Lancaster, PA.

Levitin, D. J., Chordia, P., & Menon, V. (2012). Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proceedings of the National Academy of Sciences*.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into v. *BioScience*, *51*(5), 341–352.

Lloyd, D., & Stupfel, M. (1991). The Occurance and Function of Ultridian Rhythms. *Biological Reviews*, *66*(3), 275–299. doi:10.1111/j.1469-185X.1991.tb01143.x

Mackey, M. C., & Glass, L. (1977). Oscillation and chaos in physiological control systems. *Science*, *197*(4300), 287–289.

Mackworth, N. H. (1948). The breakdown of vigilance durning prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*(1), 6–21. doi:10.1080/17470214808416738

Madurski, C., & LeBel, E. P. (2014). Making Sense of the Noise: Two Unsuccessful High-Powered Replication Attempts of Correll's (2008) 1/f Noise Racial Bias Effect. *Personality and Social Psychology, Forthcoming*.

Mandelbrot, B. B. (1967). How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*, *156*(3775), 636–638. doi:10.2307/1721427

Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, *10*(4), 422–437.

Manseau, C., & Broughton, R. J. (1984). Bilaterally Synchronous Ultradian EEG Rhythms In Awake Adult Humans. *Psychophysiology*, *21*(3), 265–273. doi:10.1111/j.1469-8986.1984.tb02933.x

Mascelli, J. V. (1998). *The five C's of cinematography: motion picture filming techniques*. Silman-James Press.

Mauldin, B. (1950). The care and handling of a heritage: One of the "scared-rabbit" generation reassures wild-eyed elders about future. *Life Magazine*, 96–100.

Méliès, G. (1899). *Cendrillon*. France: Star Film Company.

Mendelbrot, B. B. (1965). Une classe de processus stochastiques homothétiques à soi; application à la loi climatologique de HE Hurst. *Comptes Rendus Acad. Sci. Paris*, *240*, 3274–3277.

Miles, W. R. (1953). Effectiveness of Red Light on Dark Adaptation. *Journal of the Optical Society of America*, *43*(6), 435–441. doi:10.1364/JOSA.43.000435

Miles, W. R., Goodwin, C. J., & Royer, L. (2010). *Walter Miles and His 1920 Grand Tour of European Physiology and Psychology Laboratories A Reproduction of* *the Original Typescript, with a Foreword by David B. Baker and an Introduction by C. James Goodwin*. University of Akron.

Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, *114*(3), 806.

Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. *Psychological Science*, *24*(5), 776–781.

Musha, T., & Yamamoto, M. (1997). 1/f fluctuations in biological systems. *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*. doi:10.1109/IEMBS.1997.756890

Nagatani, T. (1993). Power-Law Distribution and 1/f Noise of Waiting Time near Traffic-Jam Threshold. *Journal of the Physical Society of Japan*, *62*(8), 2533–2536. doi:10.1143/JPSJ.62.2533

Neisser, U. (1967). *Cognitive psychology* (1st ed.). New York: Appleton-Century.

Neisser, U. (1998). *The rising curve: Long-term gains in IQ and related measures.* (U. E. Neisser, Ed.) (1st ed.). Washington, DC: American Psychological Association.

Osgood, B. (2007). EE261 - The Fourier Transform and its Applications. Stanford University - Electrical Engineering Department.

Peng, C., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *5*(1), 82–87.

Pilgram, B., & Kaplan, D. T. (1998). A comparison of estimators for 1f noise. *Physica D: Nonlinear Phenomena*, *114*(1–2), 108–122. doi:http://dx.doi.org/10.1016/S0167-2789(97)00188-7

Pross, N., Demazières, A., Girard, N., Barnouin, R., Santoro, F., Chevillotte, E., … Le Bellego, L. (2013). Influence of progressive fluid restriction on mood and physiological markers of dehydration in women. *British Journal of Nutrition*, *109*(02), 313–321. doi:10.1017/S0007114512001080

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. doi:http://www.R-project.org/.

Salt, B. (1974). Statistical style analysis of motion pictures. *Film Quarterly*, 13–22.

Salt, B. (1992). *Film style and technology: history and analysis*. Starword.

Salt, B. (2006). *Moving Into Pictures: More on Film History, Style, and Analysis*. Starword.

Sartwell, P. E. (1950). THE DISTRIBUTION OF INCUBATION PERIODS OF INFECTIOUS DISEASE. *American Journal of Epidemiology*, *51*(3), 310–318.

Sauter, C., Danker-Hopfe, H., Loretz, E., Zeitlhofer, J., Geisler, P., & Popp, R. (2013, June 1). The assessment of vigilance: normative data on the Siesta sustained attention test. *Sleep Medicine*. Elsevier Science,.

Schlosberg, H., & Heineman, C. (1950). The relationship between two measures of response strength. *Journal of Experimental Psychology*, *40*(2), 235–247. doi:10.1037/h0062267

Schwan, S., & Ildirar, S. (2010). Watching Film for the First Time How Adult Viewers Interpret Perceptual Discontinuities in Film. *Psychological Science*, *21*(7), 970–976.

Sedlmeier, P., Eberth, J., Schwarz, M., Zimmermann, D., Haarig, F., Jaeger, S., & Kunze, S. (2012). The psychological effects of meditation: A meta-analysis. *Psychological Bulletin*, *138*(6), 1139.

Seneta, E. W. (2010). Harold Edwin Hurst (version 4). *StatProb: The Encyclopedia Sponsored by Statistics and Probability Soceities*. Retrieved from http://statprob.com/encyclopedia/HaroldEdwinHURST.html

Shimotsu, K., & Phillips, P. C. B. (2000). *Modified local Whittle estimation of the memory parameter in the nonstationary Case*. Cowles Foundation for Research in Economics.

Smeaton, A. F., Over, P., & Doherty, A. R. (2010). Video shot boundary detection: Seven years of TRECVid activity. *Computer Vision and Image Understanding*, *114*(4), 411–418.

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, *6*(1), 1–27.

Smith, T. J., & Henderson, J. M. (2008). Edit blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, *2*(2), 6.

Stephen, D. G., Stepp, N., Dixon, J. A., & Turvey, M. T. (2008). Strong anticipation: sensitivity to long-range correlations in synchronization behavior. *Physica A:*

Statistical Mechanics and Its Applications, *387*(21), 5271–5278. doi:10.1016/j.physa.2008.05.015

Stoelting, C. H. (1930). The Dunlop Chronoscope. In *Apparatus, Tests and Supplies for Psychology, Psychometry, Psychotechnology, Psychiatry, Neurology, Anthropology, Phonetics, Physiology, and Pharmacology [Catalog]* (p. 83). Wood Dale, IL: C.H. Stolting Co.

Sundet, J. M., Borren, I., & Tambs, K. (2008). The Flynn effect is partly caused by changing fertility patterns. *Intelligence*, *36*(3), 183–191. doi:10.1016/j.intell.2007.04.002

Szendro, P., Vincze, G., & Szasz, A. (2001). Pink-noise behaviour of biosystems. *European Biophysics Journal*, *30*(3), 227–231. doi:10.1007/s002490100143

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

The Mathworks, I. (2014). MATLAB - The Language of Technical Computing. Natic, Massachusetts: The Mathworks, Inc. Retrieved from www.mathworks.com

The Mathworks, I., & The Mathworks Inc. (2007). MATLAB - The Language of Technical Computing. Natic, Massachusetts: The Mathworks, Inc. Retrieved from www.mathworks.com

The Mathworks Inc. (2007). MATLAB - The Language of Technical Computing.

Thilo, K. V, & Walsh, V. (2002). Vision: When The Clock Appears to Stop. *Current Biology*, *12*(4), 135–137. doi:10.1016/S0960-9822(02)00707-8

Thorndike, R. L. (1975). Mr. Binet's Test 70 Years Later. *Educational Researcher*, *4*(5), 3–7. doi:10.3102/0013189X004005003

Titchener, E. (1908). *Lectures on the elementary psychology of feeling and attention,*. New York: The Macmillan Company 1908.

Torre, K., Varlet, M., & Marmelat, V. (2013). Predicting the biological variability of environmental rhythms: weak or strong anticipation for sensorimotor synchronization? *Brain and Cognition*, *83*(3), 342–50. doi:10.1016/j.bandc.2013.10.002

Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *The American Journal of Psychology*, 206–219.

Tsivian, Y., & Civjans, G. (2006). Cinemetrics. *cinemetrics.lv*. Retrieved from www.cinemetrics.lv

Turner, J. F. (2004). *Frank Sinatra*. Lanham, MD: Taylor Trade Pub.

Ulrich, R., & Miller, J. (1993). Information Processing Models Generating Lognormally Distributed Reaction Times. *Journal of Mathematical Psychology*, *37*(4), 513–525. doi:10.1006/jmps.1993.1032

United Press. (1938). No TitleThey Still Know It's Time to Eat in New 28 Hour Day. *The Milwaukee Journal*, p. 1. Miilwaukee.

US National Transportation Safety Board. (2010). *Loss of control on approach, Colgan Air, Inc. operating as Continental Connection Flight 3407, Bombardier DHC-8-400, N200WQ Clarence Center, New York, February 12, 2009*. Washington, D.C.: National Transportation Safety Board.

Valentine, E. R., & Sweet, P. L. G. (1999). Meditation and attention: A comparison of the effects of concentrative and mindfulness meditation on sustained attention. *Mental Health, Religion & Culture*, *2*(1), 59–70. doi:10.1080/13674679908406332

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: an overview. *NeuroImage*, *80*, 62–79. doi:10.1016/j.neuroimage.2013.05.041

Van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 359–366.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(3), 331.

Verplanck, W. S., Collier, G. H., & Cotton, J. W. (1952). Nonindependence of successive responses in measurements of the visual threshold. *Journal of Experimental Psychology*, *44*(4).

Verplanck, W. S., Cotton, J. W., & Collier, G. H. (1953). Previous training as a determinant of response dependency at the threshold. *Journal of Experimental Psychology*, *46*(1).

Volkmann, F. C. (1986). Human visual suppression. *Vision Research*, *26*(9), 1401–1416. doi:10.1016/0042-6989(86)90164-1

Voss, R. F., & Clarke, J. (1975). “1/f noise” in music and speech. *Nature*, *258*(5533), 317–318. doi:dx.doi.org/10.1038/258317a0

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/fα noise in human cognition. *Psychonomic Bulletin & Review*, *11*(4), 579–615. doi:10.3758/BF03196615

Wald, A. (1980). *A Reprint of“ A Method of Estimating Plane Vulnerability Based on Damage of Survivors” by Abraham Wald*. Operations Evaluation Group, Center for Naval Analyses.

Wald, A., & Wolfowitz, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*, *14*(4), 378–388.

Wang, A. (2003). An Industrial Strength Audio Search Algorithm. In *ISMIR* (pp. 7–13).

Wang, A. (2006). The Shazam music recognition service. *Communications of the ACM*, *49*(8), 44–48.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of `small-world’ networks. *Nature*, *393*(6684), 440–442. Retrieved from http://dx.doi.org/10.1038/30918

Wehr, T. A. (1992). A brain-warming function for REM sleep. *Neuroscience & Biobehavioral Reviews*, *16*(3), 379–397.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The Structure of the Nervous System of the Nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *314*(1165), 1–340. doi:10.1098/rstb.1986.0056

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv För Matematik*, *2*(5), 423–434. doi:10.1007/BF02590998

Williams, C. B. (1940). A note on the statistical analysis of sentence-length as a criterion of literary style. *Biometrika*, 356–361.

Wolf, R. (1848). Nachrichten von der Sternwarte in Berne. Sonnenflacken Beobachten. *Mittheilungen Der Naturforschenden Gesellschaft in Bern*, 169–173.

Worth, S., & Adair, J. (1970). Navajo filmmakers. *American Anthropologist*, *72*(1), 9–34.

Yule, G. U. (1921). On the time-correlation problem, with especial reference to the variate-difference correlation method. *Journal of the Royal Statistical Society*, 497–537.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between Time-Series?--a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 1–63.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 267–298.