Physiology and Cognition: Reproducible Code for Data Analysis

by Raunak Pednekar

Harvey Mudd College

Advisor: Paul Zak

Abstract

Heart Rate and Electrodermal Activity data was collected while 74 volunteers watched 4 movie trailers/advertisements on one of 2 platforms (TV screen or mobile screen). Instruments collecting this data outputted it in the form of multiple excel files, making analysis cumbersome and error-prone. In this paper, I outline reproducible code that would collate this data and make analysis easier. In addition, I outline algorithms to compute metrics from this data (phasic EDA, heart rate variability) and perform cluster, regression, and time-series (ARIMA model) analyses. Key results include (1) Heart rate variability is not significantly different across different personality clusters (2) ARIMA time-series models are not ideal for analyzing physiological data.

CONTENTS

INTRODUCTION

In 1629, Rene Descartes proposed the famous mind-body problem. The mind-body problem posited that the existence of the mind is the only thing we can know with complete certainty; the body (and the physical world in general) may well be illusory. If you happen to be a descartian dualist, you may have to suspend your dualism for the purposes of this paper. This paper rests on the assumption – one that has been substantiated with plentiful evidence – that "cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capacities that are inseparably linked and that together form the matrix within which reasoning, memory, emotion, language, and all other aspects of life are meshed" (Thelen, Schöner, Scheier, & Smith, 2001, p. 1). It implies, put more simply, that we can make inferences about cognitive processes by correlating them with bodily reactions.

In this paper, we will analyze physiological responses to understand the cognitive and emotional responses that are triggered by media advertisements. Before we do so, it is useful to have an overview of modern theories about cognition and emotion. One lens of human cognition that has proven to be useful for media psychologists, is the lens of motivated attention. Motivated attention claims that the amount of attention paid to a stimulus is modulated by the stimulus' emotional/motivational significance (Lang, Bradley & Cuthbert, 1997). This is highly relevant for media psychology research because we often respond to things of motivational importance in media – especially immersive media like HDTV's or video games – as if they were really there, even though we are aware on some level that they are merely constantly recombining pixels (Reeves & Naas, 1996). This likely has to do with our long evolutionary history in which physical objects necessarily came along with their images. Another influential

model in media psychology is the Limited Capacity Model of Motivated, Mediated, Message Processing (LC4MP) (Lang, 2009). This theory states that the human mind has too limited a capacity to fully process all bits of information in a media message. A corollary to the LC4MP relevant to this paper provides greater resolution on one of the foundational questions of the field: is attention allocated consciously & voluntarily or unconsciously & involuntarily. The answer, is of course, both, and Lang, Potter & Bolls (1999) and Schneider, Dumais & Shiffrin (1984) explore the conditions for which the range from complete automaticity to entirely controlled is determined. One mechanism through which resources are automatically allocated to encoding media content, is the orienting response (also called the "what is it response"). Several features of media have been found to evoke the orienting response (measured by changes in physiology). Examples include cuts in TV (Lang, 1990; Pottter, Lang & Bolls, 2008), sound effects and voice changes in audio (Potter, 2000; Lang and Bolls, 2008) and animation in web pages (Diao & Sundar, 2004; Lang et al. 2002). In response to the orienting response that results in a temporary (phasic) physiological reaction, conscious allocation of cognitive resources result in more enduring (tonic) responses. Significant research has been done in attempting to extricate tonic and phasic responses in physiology from each other, and we will explore these concept in more detail later. Before we do so, however, let us move on from human cognition to human emotion.

It is useful, at the outset to differentiate emotion from attitudes and moods. Emotions can be conceptualized as relatively fleeting valenced reactions evoked by encountering specific stimuli (Frijda, 1994; Larsen et al 2008). Attitudes are effective reactions that are more enduring (Eagy & Chaiken, 1993), while moods -- while also more enduring that emotions -- tend to be more diffuse rather than being aimed at a particular stimulus (Frijida, 1994). It is important to

note here that emotion and cognition are not completely separable processes. Tucker, Derrybrry, and Luu (2000) showed that emotional processes are implemented in an interconnected network that spans the sub-cortical areas of the brain (traditionally emotional) and the cortical areas (traditionally cognitive). The brain areas that makeup this emotional network can be thought as being connected in both a top-down and bottom-up fashion. As Cacioppo, Gardner, and Bernston (1999) put it, cognition shapes emotion and emotion helps construct cognition. In 2007, Peter Lang and Margaret Bradley noted three ways in which emotion is expressed – verbal, behavioral, and physiological. Much emotional and cognitive processing is hidden below explicit behavioral and verbal responses. It likely occurs in the physiological activity of the central and peripheral nervous system. It is this activity that we will attempt to analyze here.

In this paper, we will analyze heart rate and electrodermal activity (EDA). To understand how these two physiological measures can shed insight on "the matrix within which reasoning, memory, emotion, language, and all other aspects of life are meshed", lets first take a short tour of the nervous system. Nerve cells can be functionally grouped into sensory and motor cells. Motor neurons can be further subdivided based on whether they serve skeletal muscles or organs and glands. The cells that serve organs and glands make up the Autonomic Nervous System (ANS) which controls the body's unconscious actions. Both heart rate and electrodermal activity depend on the ANS. The ANS can be further subdivided into the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS). The SNS is responsible for preparing the body for activities vital for procreation and survival, or as Bear, Connors & Paradiso (2007) put it, "the four F's: fight, flight, fright and sex". So, when you are anticipating something that scares you – an upcoming thesis presentation in front of all your professors, for example, your SNS will activate resulting in increased lung & heart activity (Kalat, 2007). The Parasympathetic Nervous

System (PNS) is the branch of the ANS has been described as being devoted to "rest, repair, and enjoyment" (Andreass, 2000, p. 67) or that "facilitates vegetative non-emergency responses by organs" (Kalat, 2007).

The heart rate (RR) and electrodermal activity (EDA) data that we will analyze in this paper was collected while 74 participants watched 4 movie trailers/advertisements (*Fault in Our Stars, Entourage, Get Hard, Victoria's Secret*) on 2 platforms (computer monitor and mobile phone) in a within-subject study. In addition to collecting physiological data, participants were also asked to fill a pre-experiment and post-experiment survey. The pre-experiment survey included questions about the participant's personalities and moods, while the post-experiment survey attempted to get at how participants felt about the movie trailer/advertisements. **The goal of this paper is to design reproducible methods to efficiently analyze this physiological and survey data.**

Before we go on to data analysis methods, let us first learn a little more about heart rate & electrodermal activity to better understand stand how they relate to cognitive and emotional processing.

HEART RATE

The heart, like most organs, is innervated by the sympathetic and parasympathetic branches of the autonomic nervous system. A useful rule of thumb is that the sympathetic system ("flight, fight, fright, sex") makes the heart speed up, while the parasympathetic nervous system -- when engaged to provide greater cognitive resources to the stimulus -- makes it slow down. Annoyingly, stimuli like the movie trailers used in our experiments activate both the sympathetic and parasympathetic nervous system. This is because increases in interest and

attention often go along with increased emotional responses to stimulus. As Reeves and Nass (1996) demonstrated, our brains don't recognize arousing media as being mediated but think of them as something that is real and present.

While it is true that the sympathetic system increases heart rate, media is unlikely to be so arousing that sympathetic innervation will completely overcome the deceleration from the parasympathetic system. Multiple previous studies have as people allocate more cognitive resources to a television stimulus, heart rate decreases (Lang, 1990; Lang, 1994; Lang, Bolls et al., 1999; Lang et at., 1996). This decrease in heart rate has been generally interpreted as a result of "individuals allocating more cognitive resources to encoding information from the environment" (Potter, 2012).

Because of the opposing effects of the parasympathetic and sympathetic nervous system on heart rate, some researchers (Rajava, 2004) have suggested using heart rate variability (HRV) to tease out the individual contributions of these two branches of the autonomic nervous system. HRV has been described as "a measure of the continuous interplay between sympathetic and parasympathetic influences on heart rate that yields information about autonomic flexibility and thereby represents the capacity for regulated emotional responding" (Appelhans, 2006).

ELECTRODERMAL ACTIVITY (EDA)

In the late 19th century, a French neurologist named Charles Fere observed that there were systematic changes in the electrical activity of skin (Andreassi, 2007). Fere also discovered that the amount of conductivity could be changed by external emotional physical or emotional stimuli. Changes in electrical conductivity come from changes in levels of sweat on skin (Sweat is salty- hence conducts electricity). There are two kinds of sweat glands underneath the outer

layer of human skin – apocrine and eccrine. Apocrine sweat glands are responsible for regulating body temperature and are thus not of much psychological importance. Eccrine sweat glands, in contrast, are of significant psychological importance. For instance, eccrine sweat glands on feet soles and hand palms are believed to produce sweat when faced by fight or flight situations. This is supposed to confer the adaptive advantage of making grasping easier (Edelberg, 1972). **Unlike the heart and most other glands & organs, eccrine sweat glands are entirely innervated by the sympathetic branch of the ANS (Boucsein, 1992)**. Oddly enough, the neurotransmitter that activates eccrine glands (acetylcholine) is usually involved in parasympathetic activation (Shields et al. 1987). Many stimuli that evoke emotion have been demonstrated to evoke detectable changes in skin conductance; for example – music (Grewe et al. 2007), pictures (Lang et al, 1993), and emotional film (Codispoti et al 2008). A reliable association between sympathetic activity and changes in skin conductance as far back as 1987 (Shields et al).

While skin conductance level measurements have a rich history in psychology, EDA measurements are prone to artifacts from unimportant experimental conditions. The most important of these is movement. Changes in skin conductance level can be elicited by pressure or stretching of the skin at the recording site. Burbank and Webster (1978) found quantitative relationships between skin stretching at the volar side of the forearm and elicited EDA artifacts. Temperature too has been shown to affect SCL. Conklin (1951) investigated the relationship between three different room temperatures (21.9, 26.9, and 29. C) and the SCL measured at three different sites (wrist, forehead, and palm). He observed a positive correlation between temperature and SCL. Just like with heart rate and other physiological measures, demographic differences have been found as well. Females in general display higher SCLs, while males tend to show a greater electrodermal reactivity under conditions of stimulation (Boucsin, 2007).

African-Americans tend to have higher SRLs or lower SCLs than Caucasians under resting conditions (Boucsin, 2007)

PHASIC VS TONIC RESPONSES

The last concept to understand before we finally go ahead to the data analysis, is the difference between tonic and phasic responses. Tonic responses occur over a comparatively long time period, and are in response to the general experimental condition (as opposed to a specific stimulus) . Phasic responses are far shorter lived, and are in reaction to a specific stimulus.

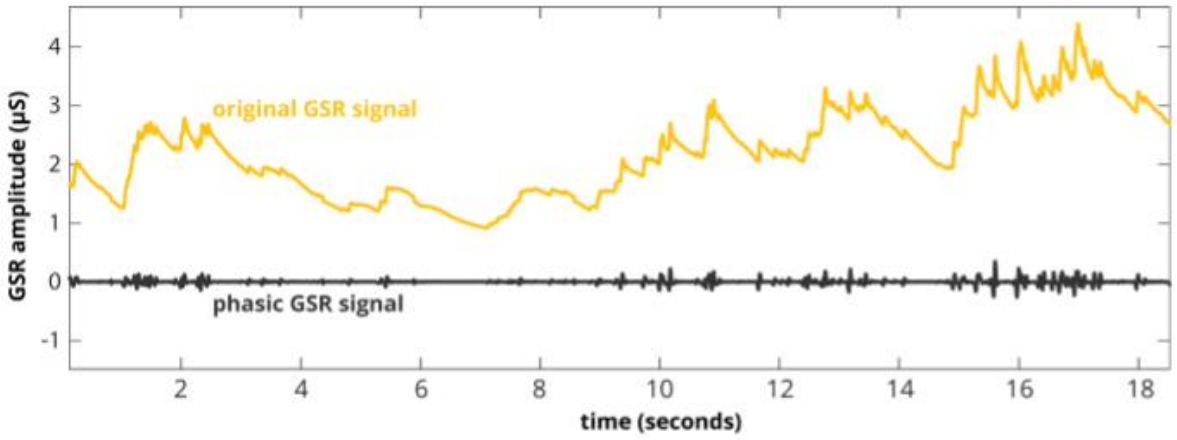


FIGURE 1: PHASIC VS TONIC COMPONENTS

REPRODUCIBLE CODE TO CLEAN AND ORGANIZE DATA

INTRODUCTION

The instruments used to collect heart rate and skin conductance data store output in multiple csv files. One csv file is created per experimental condition per physiological metric. As

there are 4 movies, 2 platforms, and 2 physiological metrics, this means that heart rate and skin conductance data from this experiment was stored in 18 (4×2×2 + 2 baseline) csv files (Each file contains data from all participants for that condition). This makes data analysis cumbersome and error prone. Figure 1 below is a screenshot of what the typical monitor of a researcher would look like as they were analyzing this data (The columns represent individual participants).

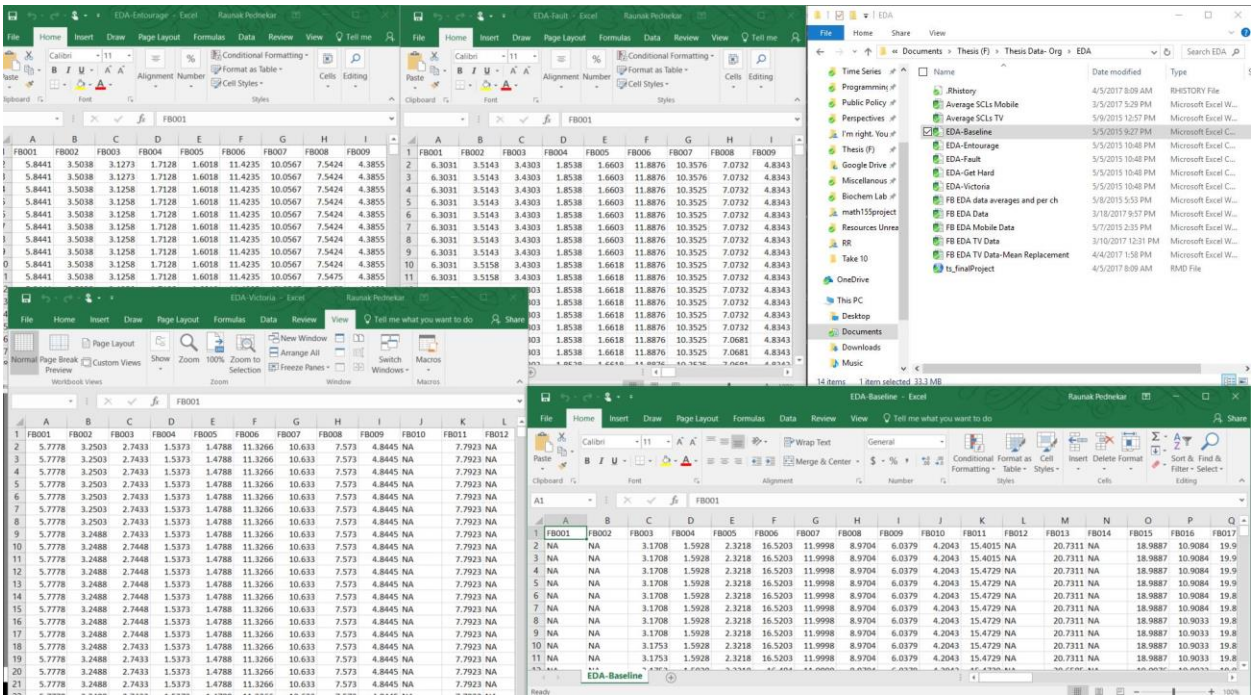


FIGURE 2: A SCREENSHOT OF WHAT THE TYPICAL RESEARCHERS MONITOR MIGHT LOOK LIKE WHEN ANALYZING PHYSIOLOGICAL DATA. THE NUMBER OF CSV FILES MAKES CROSS-CONDITION ANALYSIS CUMBERSOME AND ERROR-PRONE

Some researchers in the lab would copy and paste columns into one single excel file. Since movie and platform information is encoded, not in the cells themselves, but in the *names* of excel files, this method, apart from being time consuming is dangerously prone to error. My

first task was to write reproducible code to organize this data to make analysis easier and less error-prone. This task had to achieve two objects:

(1) Ensure the code was easily adaptable to output from other experiments. I wanted to write code, that, with small changes, could be used to organize the messy output from future lab experiments.

(2) Ensure all information was encoded *within* the output files (as opposed to in the titles as is currently the case).

RESULTS

A snippet of the final output file is illustrated in figure _ below.

B	C	D	E	F
Person	Platform	Movie	RR	EDA
1	1	1	1.210754	5.8441
1	1	1	1.207491	5.8288
1	1	1	1.183268	5.8186
1	1	1	1.130217	5.8084
1	1	1	1.177156	5.7982
1	1	1	1.062025	5.788
1	1	1	1.111161	5.7829
1	1	1	1.061751	5.7727
1	1	1	0.94995	5.7625
1	1	1	0.986873	5.7472
1	1	1	1.128771	5.7472
1	1	1	1.115248	5.7421
1	1	1	1.193338	5.7217
1	1	1	1.269519	5.7115
1	1	1	1.257085	5.7013
1	1	1	1.253979	5.6962
1	1	1	1.214357	5.6911
1	1	1	1.163238	5.6809
1	1	1	1.14386	5.6656
1	1	1	1.127688	5.6605
1	1	1	1.099836	5.6452
1	1	1	1.170607	5.6401
1	1	1	1.111005	5.6401
1	1	1	1.11965	5.6299
1	1	1	1.118086	5.6248

FIGURE 3: SNIPPET OF FINAL CLEANED UP DATA

As you can see, the entirety of the experimental condition is represented in the file itself. The person, platform, and movie numbers are stored in cells. We will look at the advantages of doing this later. Before that, let us look at objective 1- Make code easily adaptable to output from other experiments – was achieved. Figure _ below shows how the details of this particular experiment were encoded

```
# Global Constants and Dictionaries
MOVIE_NAME_DICT = {'Entourage': 1, 'Fault': 2, 'Get Hard': 3, 'Victoria': 4}
PLATFORM_DICT = {'TV': 1, 'Mobile': 2}

# Create dataframe: Person, Platform, Movie, RR, EDA
dataframe_columns = ['Person', 'Platform', 'Movie', 'RR', 'EDA']
```

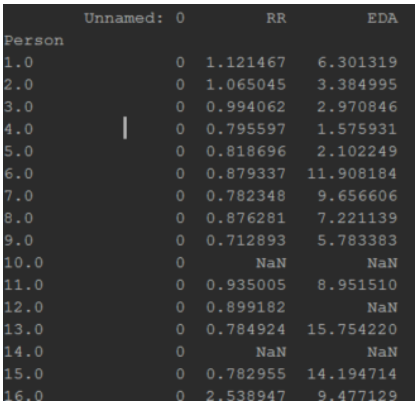
FIGURE 4: ADAPTABLE PART O DATA CLEANING CODE. THE APPROPRIATE LINES HAVE TO BE CHANGED IN ORDER FOR THIS CODE TO BE DEPLOYED FOR FUTURE SITUATIONS

The line starting with “MOVIE_NAME_DICT” encodes the relationship between movies and code numbers, while “PLATFORM_DICT” does the same for different platforms. “dataframe_columns” sets column header for the final csv output in figure _. In future experiments with different/more/less experimental conditions, these 3 lines would have to be changed to produce the output in figure 1.

The advantages of storing data cleanly are numerous. Apart from making data analysis less error prone, it also makes data analysis easier. For instance, calculating average EDA and heart rate values for all participants across movies and platforms can be achieved with a single line of code.

```
person_grouped_data = dataframe.groupby(["Person"]).mean()
```

The output from this line is in figure _ below. (EDA values can be similarly averaged across platform or movies if so desired.)



Person	Unnamed: 0	RR	EDA
1.0	0	1.121467	6.301319
2.0	0	1.065045	3.384995
3.0	0	0.994062	2.970846
4.0	0	0.795597	1.575931
5.0	0	0.818696	2.102249
6.0	0	0.879337	11.908184
7.0	0	0.782348	9.656606
8.0	0	0.876281	7.221139
9.0	0	0.712893	5.783383
10.0	0	NaN	NaN
11.0	0	0.935005	8.951510
12.0	0	0.899182	NaN
13.0	0	0.784924	15.754220
14.0	0	NaN	NaN
15.0	0	0.782955	14.194714
16.0	0	2.538947	9.477129

FIGURE 5: EDA AND RR MEANS ACROSS MOVIE PLATFORM COMBINATIONS. THIS TASK WOULD TAKE 20+ FUNCTION CALLS BEFORE. NOW, IT TAKES A SINGLE LINE OF CODE TO PRODUCE

Note that this task would have previously required one to delve into 16 different excel files, calculate mean value for each column, then average these values across the 16 different files.

Means (or medians or standard deviation, quartile ranges or similar statistics) can also be calculated across combinations of experimental conditions. (per movie per participant, for example)

```
personMovie_grouped_data = dataframe.groupby(["Person", "Movie"]).mean()
```

Person	Movie	Unnamed: 0	RR	EDA
1.0	1.0	0	1.131192	5.925808
	2.0	0	1.100899	7.363036
	3.0	0	1.147243	6.497347
	4.0	0	1.097994	6.554873
2.0	1.0	0	1.065196	3.377970
	2.0	0	1.069144	3.706300
	3.0	0	1.114753	3.577353
	4.0	0	1.041947	3.166710
3.0	1.0	0	0.974669	2.922865
	2.0	0	1.024593	3.673061
	3.0	0	1.024165	3.362800
	4.0	0	0.995989	2.578710
4.0	1.0	0	0.776377	1.568249
	2.0	0	0.817640	1.761080
	3.0	0	0.784666	1.688458
	4.0	0	0.811429	1.457466
5.0	1.0	0	0.811611	2.070049
	2.0	0	0.841734	2.691741
	3.0	0	0.842463	2.459432

FIGURE 6: FIGURE 4: EDA AND RR MEANS ACROSS MOVIE-PERSON COMBINATIONS. THIS TASK WOULD TAKE 15+ FUNCTION CALLS BEFORE. NOW, IT TAKES A SINGLE LINE OF CODE TO PRODUCE

Note that all this data can easily be written to a csv file that would open in excel

METHODOLOGY

The two functions that do the heavy lifting for this code contain nested for loops. The first set loops through individual files. As mentioned before, there is one csv file for each movie-platform combination.

```
# For each combination of movie and platform
for movie in MOVIE_NAME_DICT:
    for platform in PLATFORM_DICT:
        # populateDataframe(df,movie,platform)
        df = populateDataframe(df, movie, platform)
```

FIGURE 7: LOOPING THROUGH ALL FILES

The second function loops through every column in each file. This is because each individual column contains data for one participant.

```
for cont_col
    # Loop through index
    for i in range(min_index):
        index_EDA = index_name_EDA[i]
        index_RR = index_name_RR[i]
        RR_col = RR_data[col][index_RR]
        df2 = pd.DataFrame([[int(col[-3:]), # Column to participant number
                            PLATFORM_DICT[platform],
                            MOVIE_NAME_DICT[movie_name],
                            RR_col,
                            EDA_data[col][index_EDA]]],
                            columns = dataframe_columns)
        df = df.append(df2)
return df
```

FIGURE 8:LOPPING THROUGH ALL COLUMNS IN ONE FILE

In addition, there is a function to downs-ample EDA data by a factor of 256, to make sure it is at the same frequency as heart rate. There is also a function to write the output table to a csv file openable in excel.

PERSONALITY CLUSTERING AND HEART RATE VARIABILITY

THEORY

The pre-experiment survey contained demographic questions (age/sex/race etc.), mood questions ("The following questions contain a list of words that describe feelings. Read each word, and rate, on a scale of 1-5 how strongly it applies to your current mood?" – jittery/anxious/excited etc.), and personality questions ("How well do the following statements describe your personality- I see myself as generally trusting/happy/lazy etc.). In this part of the paper we discuss how we clustered people based on personality questions and analyzed the

dependence of heart rate variability on these clusters. Before we dive into results and code, let us learn about the heart rate variability and k-means clustering.

Heart Rate Variability (HRV) is a measure of variance in heart rate. It reflects the degree to which cardiac activity can be regulated to meet changing situational demands. It reflects the immediate output of the central autonomic nervous system (both the sympathetic and parasympathetic branches) and could be viewed as a proxy for an individual's ability to regulate physiological responses in the context of emotional expression (Thayer and Lane, 2000). Diminished HRV has been linked as an independent risk factor for several negative cardiovascular outcomes (Curtis & O'Keefe, 2002) and as a proxy for underlying cardiovascular disease processes (Stys & Stys, 1998). HRV has also been linked to various outcomes relevant to regulated emotional responding. Examples of these outcomes include coping mechanisms s (Fabes & Eisenberg, 1997), attention allocation (Johnsen et al., 2003), and depression (Balogh, Fitzpatrick, Hendricks, & Paige, 1993). With HRV being related to these outcomes, it would be reasonable to expect different HRV levels for people with different personality traits. To examine HRV differences among people with different personality groups, we will have to first cluster our participants by personality.

The clustering technique I used is called k-means clustering. Clustering, in general, is the grouping of data, in such a way that objects in the same group (or cluster) or more similar to each other than those in a different cluster. K-means is one popular algorithm to cluster data. It works by assigning clusters to different centroids until the sum of squares distance from each point to the centroid is minimized. It works best when 3 assumptions are met:

- k-means assume the variance of the distribution of each attribute (variable) is spherical;

- All variables have the same variance;
- Each cluster has roughly equal number of observations;

These assumptions are likely to be met because of the nature of our personality data. Each personality trait is ranked on a scale of 1-5, allowing for similar variance and sphericity. The number of people in each personality cluster is likely to be roughly equal (order of magnitude). For these reasons, I picked k-means over other clustering techniques like hierarchical clustering.

RESULTS AND METHODS

There are 6 important functions in this section of the report. (1) Algorithm to compute resting HRV (2) Picking personality dimensions to cluster on (3) Checking for correlations within personality dimensions (4) Making a scree plot to learn what the optimal number of clusters is (5) Making Cluster plot (6) Performing ANOVA for HRV on different clusters.

I calculated the Root Mean Square of the Successive Differences (RMMSD) component of heart rate variability. This is because RMSSD has been heavily studied, and as mentioned in the introduction of this section, been linked to various outcomes relevant to emotional regulation (and heart disease). RMSSD measures have been calculated at rest for each participant using the following algorithm:

```python
def hrv_calculator(rr):
    rr_diff = rr.diff()
    rr_diff_sqr = np.square(rr_diff)
    hrv_squared = rr_diff_sqr.mean()
    hrv = np.sqrt(hrv_squared)
    rmssd = hrv * 1000
    return rmssd
```

FIGURE 9: HEART RATE VARIABILITY CALCULATER. RMSD METRIC

As can be seen, interbeat interval is (1) differenced (2) squared (3) averaged and (4) multiplied by a 10000, to get RMSSD. The input to this function is a matrix (or spreadsheet) where each column contains a time-series of an individual's heart rate.

The extract variable function below allows you to input the list of variables you would use for cluster analysis.

```r
## Extracting variables for cluster analysis
extract_variables <- function(dem, cols, colnames){

  dem_imp <- dem[ , cols]
  colnames(dem_imp) <- colnames
  dem_imp <- sapply(dem_imp, as.numeric)
  return(dem_imp)

}
```

```r
c("reserved", "trusting","outgoing", "conscentious", "nervous"))
```

For my test runs I extracted 5 variables that loosely correspond to the "Big 5" personality traits.

After extracting variables, its important to make sure they aren't strongly correlated. The pairs function is useful in creating quick scatter plots to do this

FIGURE 10: CHECKING FOR CORRELATIONS IN VARIABLS CHOSEN FOR CLUSTER ANALYSIS

As there don't seem to be strong correlations, we can move on.

The scree plot function below outputs a plot that would guide the analyst as to what the optimal number of clusters is. It does this by calculating the following ratio:

$\frac{Within\ cluster\ sum\ of\ squares}{Between + Within\ cluster\ sum\ of\ squares}$. This ratio plotted against number of clusters will be a monotonically decreasing function. The trick is to identify the "elbow" of the plot. This will inform you what the appropriate number of clusters is.

```r
##Scree PLots. Plot rwithin cluster: between cluster residual ration "Elbow" in plot determines
make_scree <- function(dem_imp){


  # Initialise ratio_ss
  ratio_ss <- rep(0, 7)


  set.seed(3)
  for (k in 1:7) {

    # Apply k-means to school_result: school_km
    kmeans <- kmeans(dem_imp, k, nstart = 20)

    # Save the ratio between of WSS to TSS in kth element of ratio_ss
    ratio_ss[k] <- kmeans$tot.withinss / kmeans$totss

  }

  # Make a scree plot with type "b" and xlab "k"
  plot(ratio_ss, type = "b", xlab = "k", main = "Scree Plot")

}
```

FIGURE 11: CODE FOR MAKING SCREE PLOT

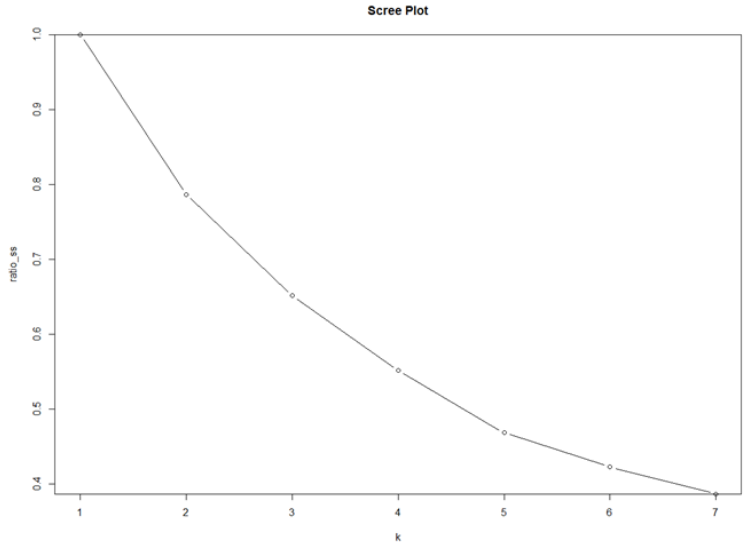


FIGURE 12: SCREE PLOT. HELPS DETERMINE OPTIMUM NUMBER OF CLUSTERS

The scree plot in my test run did not have a clear elbow. A moderate cluster size of 3 or 4 would be defensible in this situation.

What's left to do after this is plot the cluster plots. The function and plot are below.

```r
cluster_plot <- function(dem_imp , n){
  set.seed(3)
  results <- kmeans(dem_imp, n)
  plot(jitter(dem_imp$trusting), jitter(dem_imp$nervous)
  return(results$cluster)
}
```

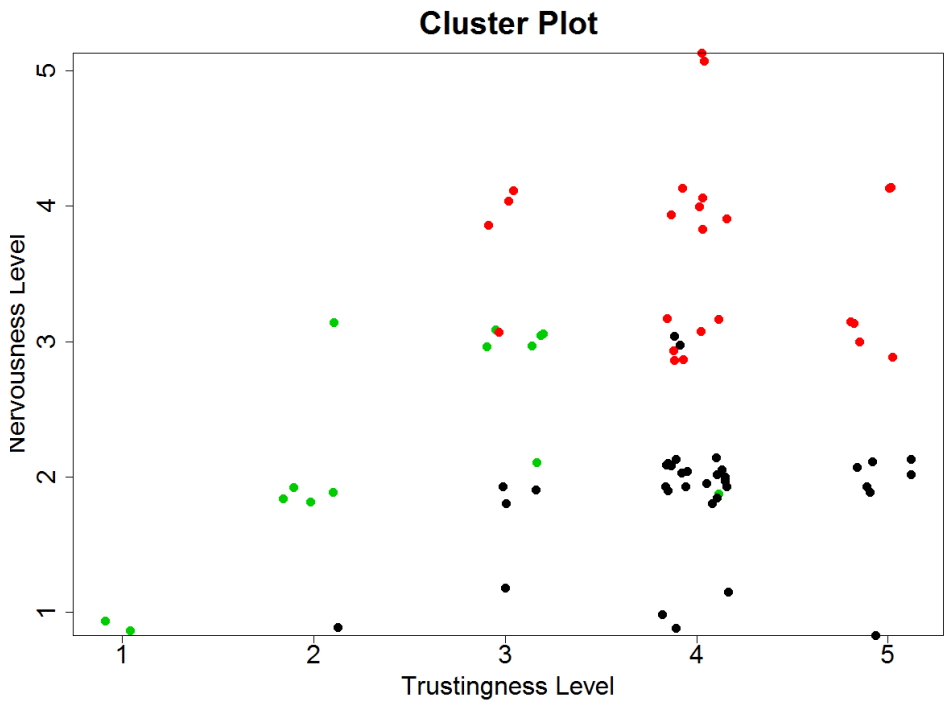


FIGURE 13: CLUSTER PLOT SHOWS 3 CLUSTERS ON NERVOUSNESS AND TRUSTING LEVEL DIMENSIONS,

As can be seen, 3 distinct clusters of people can be identified on the nervousness and trusting level dimensions.

Unfortunately, there were no statistically significant differences in the heart rate variabilities of the different personality clusters

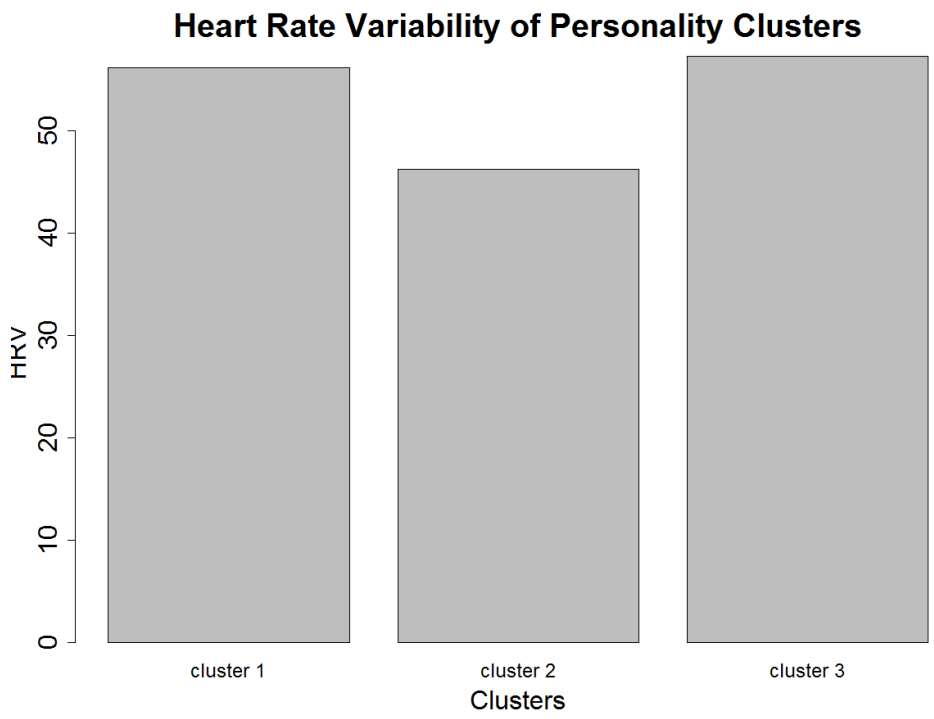


FIGURE 14: NO STATISTICALLY SIGNIFICANT DIFFERENCES IN HRV OF DIFFERENT PERSONALITY CLUSTERS

PREDICTIBILITY OF TONIC PHYSIOLOGY ON ADVERTISEMENT OUTCOMES

RESULTS

I also automated the process of plotting regression plots and extracting regression coefficients. As a test run, I plotted level of movie enjoyment against tonic EDA and tonic RR. In addition, I extracted regression coefficients for a model with movie enjoyment as the dependent variable and RR & EDA as the independent variables

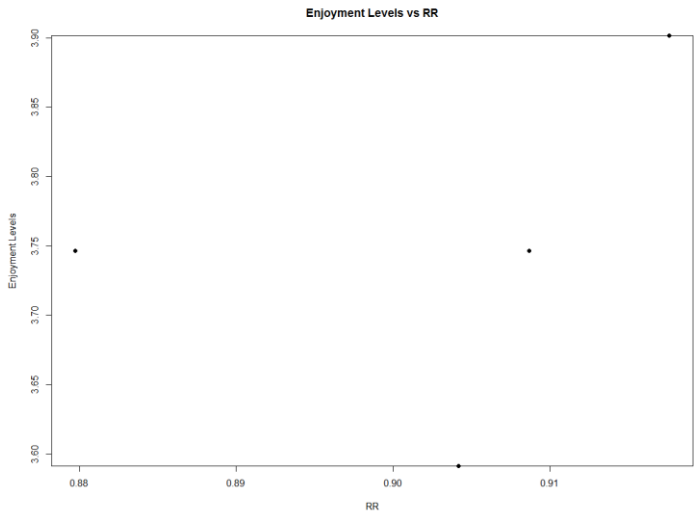


FIGURE 15: ENJOYMENT LEVEL VERSUS HEART RATE FOR 4 MOVIES. NO STATISTICALLY SIGNIFICANT RELATIONSHIP

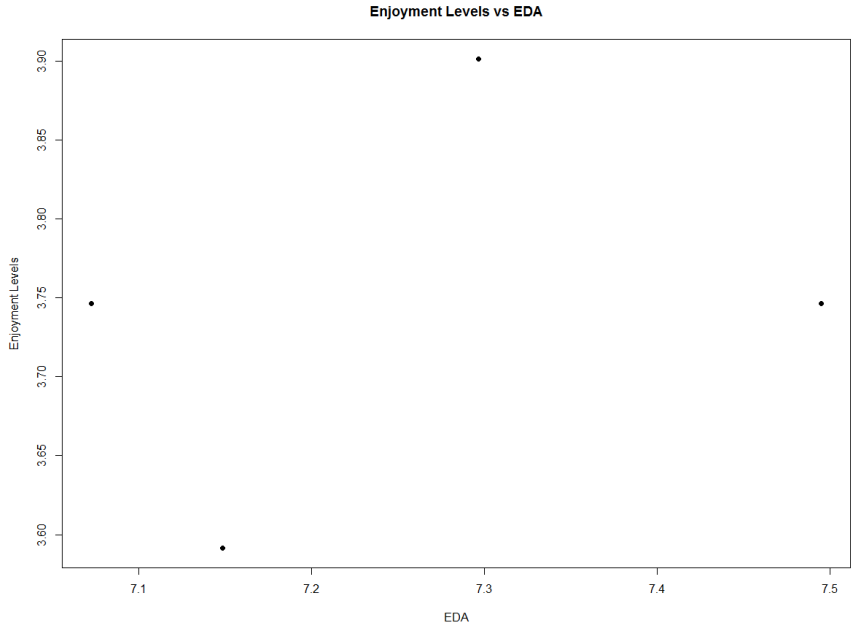


FIGURE 16: ENJOYMENT LEVEL VERSUS ELECTRODERMAL ACTIVITY FOR 4 MOVIES. NO STATISTICALLY SIGNIFICANT RELATIONSHIP

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.3269     6.5764   0.202    0.873
phys$RR       1.7074     9.8404   0.174    0.891
phys$EDA      0.1211     0.8557   0.142    0.910

Residual standard error: 0.2041 on 1 degrees of freedom
Multiple R-squared:  0.1319,    Adjusted R-squared:  -1.604
F-statistic: 0.07597 on 2 and 1 DF,  p-value: 0.9317
```

As can be seen in the plots and regression numbers, no statistically significant relationship between movie enjoyment levels and physiological metrics can be seen. This is not entirely unexpected given that heart rate and EDA track only intensity of cognitive and emotional processing, and not valence.

```
scatter_plots <- function(phys, post_imp_means)

regression_coffiecient <- function(phys, post_imp_means)
```

Note that the functions used to create the scatter plots and extract regression coefficients can be applied to other sets of independent and dependent variables by future researchers.

TIME SERIES MODELS

I am currently taking a class called Time Series analysis. The core of this class involves using ARIMA models to model time series data. A quick survey of the literature informed me that nobody had used ARIMA models to model heart rate or EDA data.

I designed an experiment to analyze whether ARIMA models could extract meaning from heart rate data. I modelled part of baseline heart rate data of one individual. If the ARIMA forecast from this model better predicted the other basline data (compared to movie data) my experiment would be successful. This is better explained visually

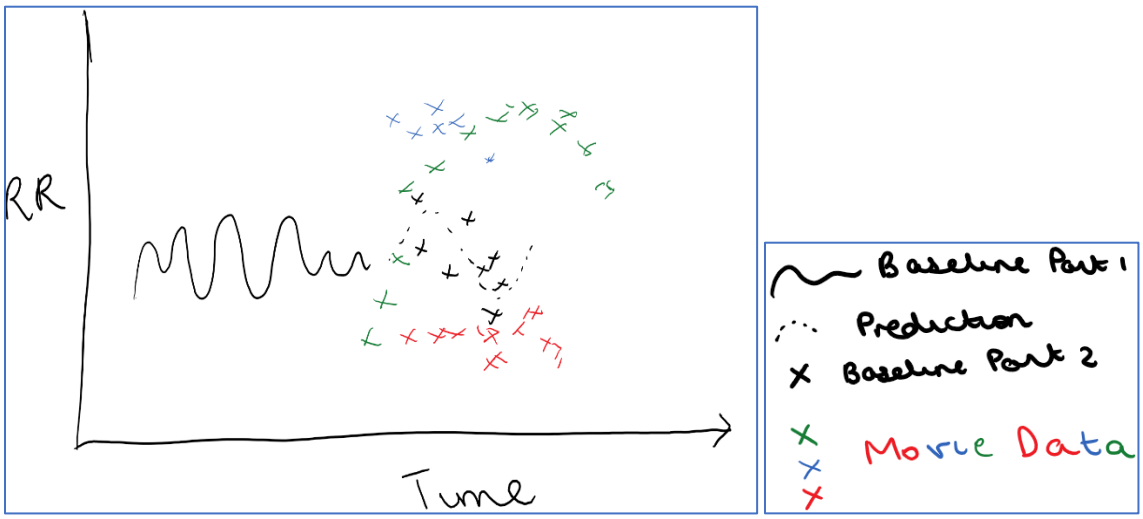
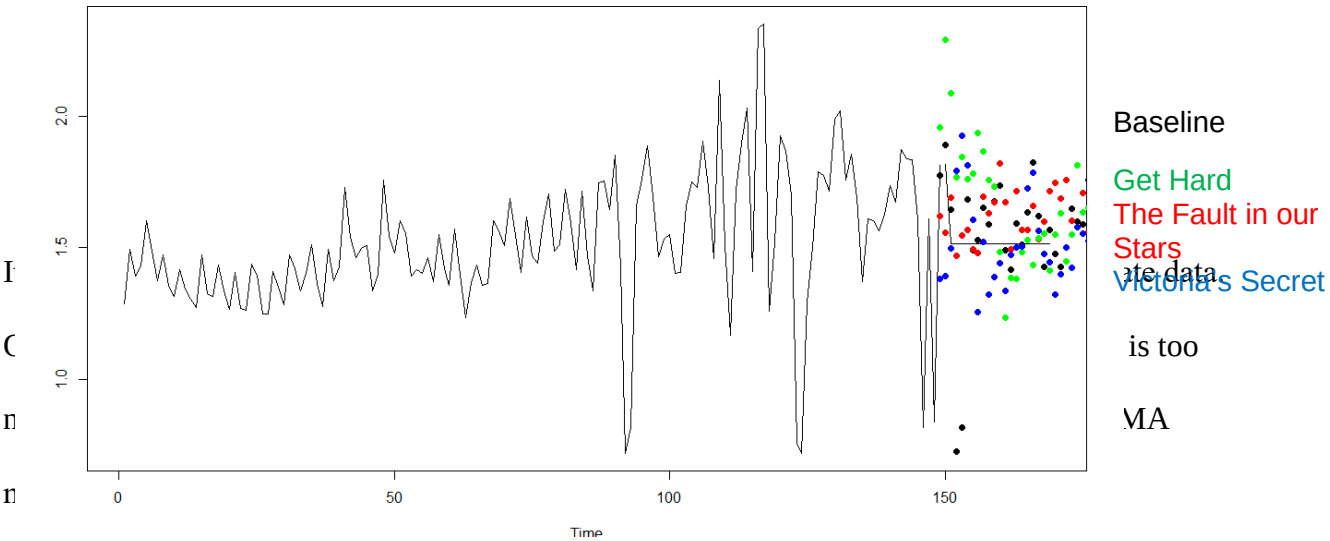


FIGURE 17: WHAT SUCESSFUL ARIMA MODELLING WOULD LOOK LIKE. PAST BASELINE HEART RATE DATA BETTER PREDICTS FUTURE BASLINE DATA THAN FUTURE MOVIE DATA. THE ARIMA MODEL HAS CAPTURED DIFFERENTIAL INFORMATION

Is this what we see? No, it isn't



I te data

C is too

n MA

n

FIGURE 18: ACTUAL ARIMA MODEL. NOT VERY USEFUL IN DISTINGUISHING RESTING FROM ENGAGED HEART RATE

ARIMA models don't seem to be appropriate

CONCLUSIONS AND FUTURE DIRECTIONS

The code I've outlined in this paper should allow for faster, easier, and less error-prone data analysis by future researchers. For instance, much of the scree plot and cluster code can be used to cluster on other personality or demographic dimensions. All of my analysis so far has been done on tonic data. I've written some code to decompose EDA data into tonic and phasic components. Phasic components of the data can be analyzed. For instance the number and magnitude of EDA peaks can be compared for different movies/genders/platforms. The process of counting and averaging the magnitude of these peaks can also be automated.

```
def make_tonic_EDA(EDA_total, n = 9):
    EDA_tonic_1 = EDA_total["FB001"].as_matrix()
    EDA_tonic_1 = np.cumsum(EDA_tonic_1)
    EDA_tonic_1[n:] = EDA_tonic_1[n:] - EDA_tonic_1[:-n]
    EDA_tonic_1 = EDA_tonic_1[n - 1:] / n
    EDA_tonic = np.zeros(EDA_tonic_1.shape[0] + n-1)
    EDA_tonic[:(n-1)/2] = EDA_tonic_1[0]
    EDA_tonic[((n - 1) / 2): -((n-1)/2)] = EDA_tonic_1
    EDA_tonic[-((n - 1) / 2):] = EDA_tonic_1[-1]

    return EDA_tonic
```

FIGURE 19: CODE TO EXTRACT TONIC AND PHASIC COMPONENTS FROM EDA DATA

ACKNOWLEDGEMENTS

- Paul Zak
  - My adviser
- Jorge Barraza
  - A second adviser, really
- Elizabeth Terris
  - For the data and explanations
- HMC Chemistry Department
  - For letting me follow my interest in thesis
- TED
  - for introducing me to Paul

REFERENCES

- Andreassi, J. L. (2000). *Psychophysiology human behavior and physiological response*. Mahwah, NJ: Erlbaum Associates.
- Andreassi, J. L. (2007). *Psychophysiology: human behavior and physiological response*. Mahwah, NJ: Lawrence Erlbaum.
- Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology,10*(3), 229-240. doi:10.1037/1089-2680.10.3.229
- Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: exploring the brain*. Philadelphia: Wolters Kluwer.
- Boucsein W. Plenum University Press; New York: 1992. Electrodermal activity

- Codispoti, M., Surcinelli, P., & Baldaro, B. (2008). Watching emotional movies: Affective reactions and gender differences. *International Journal of Psychophysiology,69*(2), 90-95. doi:10.1016/j.ijpsycho.2008.03.004
- Edelberg R (1972) Electrical acitivity of the skin: its measurement and uses in psychophysiology. in Handbook of psychophysiology (pp 367-418). New York: Holt
- Ekman, P., & Davidson, R. J. (1994). *The nature of emotion: fundamental questions*. New York: Oxford University Press. pp 59-67
- Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007). Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. Emotion, 7(4), 774-788. doi:10.1037/1528-3542.7.4.774
- Kalat, J. W. (2007). *Biological psychology*. Belmont, Calif.?: Wadsworth Cengage Learning.
- Lang, A,, Bolls, P., Potter, R. F., & Kawahara, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. lournal of Broadcasting & Electronic Media, 43(4), 451 -475.
- Lang, A. (1 990). Involuntary attention and physiological arousal evoked by structural features and mild emotion in TV commercials. Communication Research, 17(3), 275-299. 30(4), 387-41 3. 674 journal of Broadcasting & Electronic MedidDecember 2004
- Lang, A. (1994). What can heart rate tell us about thinking? In A. Lang (Ed.), Measuring psychological responses to media messages (pp. 99-1 11). Hillsdale, NJ: Erlbaum. Lang,

A. (2000). The limited capacity model of mediated message processing. journal of Communication, 50(1), 46-70.

- Potter, R. F., & Bolls, P. D. (2012). *Psychophysiological measurement and meaning: cognitive and emotional processing of media*. New York: Routledge.

- Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences,24*(1), 1-34. doi:10.1017/s0140525x01003910