The Pioneer Research Journal

An International Collection of Undergraduate-Level Research

Volume 3

2016

Pioneer® academics

The Pioneer Research Journal

An International Collection of Undergraduate-Level Research

Contents

Culture and Pain: The Effect of Culture on the Production of Endorphins in the Brain

Joshua E. Roth

Author background: Joshua E. Roth grew up in the United States and currently attends Northside College Preparatory High School, located in Chicago, Illinois. His Pioneer seminar topic was in the field of neuroscience and titled "Understanding the Sense of Touch through Neuroscience."

INTRODUCTION

Pain is well known to have a variety of different effects on the human body. It can be described in a multitude of ways, including sharp, dull, chronic, and acute. Pain information is received from sensory stimuli using specific pain nerve fibers, which are then transmitted to the brain, creating the perception of pain. One might assume that every person feels pain the same way, to the same degree, and interprets pain the same way as everyone else. However, perception of pain is not uniform or consistent across all people and it can be affected by a variety of factors. One of the most important and well-documented factors that can influence the perception of pain and its effect on it is culture. People's culture can change the way that they feel pain (Turner and Chapman, 1982, Melzack and Wall, 1983). However, even though it is well known that culture can play an important role in perceiving and managing pain, little is known about the physiological mechanism that may be initiated by cultural practices to manage or block pain. Do cultural practices and perceptions cause a person to release endorphins? Do people manage pain or perceive little pain during painful cultural practices as a result of the placebo effect? These are questions that current research cannot answer. In order to answer these questions, culture must be isolated as a factor in the perception and management of pain. Such an experiment would isolate the effects of culture by eliminating the placebo effect and endorphins by administering naloxone, which is known to block receptor sites from endorphins. By administering naloxone to native people participating in cultural practices, it can be determined if the lack of pain cause by participating in cultural practices can be caused by endorphins or other physiological changes in the person.

In order to set up the experiment, it is important to first explain the anatomy of pain and delve deeper into sensation and how sensation leads to pain. The next section will focus on the perception of pain and the ways in which it can be inhibited. This includes a discussion of naloxone and the placebo effect. The paper will then address the challenge of measuring pain and the different methods and scales that have been developed to measure it as well as the weaknesses of each. Next, the role of culture in pain perception will be explored followed by a discussion of the ways in which different cultures manage pain. The cases of the Stere-Mawe tribe of the Amazon and the indigenous people of the Deccan region in India will be presented. The perception of pain in these two groups will be studied in the experiment. Before explaining the experiment, the paper will return to a discussion of the role of culture in pain perception in order to establish the focus of the experiment. Finally, the precise methods of the experiment along with possible results will be presented.

THE ANATOMY OF PAIN

The anatomy of a neuronal cell is essential to understanding sensation and pain. The neuron (or nerve cell) is composed of several important pieces: the dendrites, the soma, and the axon. The neuronal membrane separates the spherical soma (the center of the cell) from the outside and is filled with a fluid rich in potassium called cytosol. Within the cell body of a neuron are the same organelles found in all organic cells. Everything within the cell membrane is called the cytoplasm (excluding the nucleus). The cytoskeleton of a neuron is held within the neuronal membrane and gives the neuron its characteristic three-dimensional shape (Bear, 2007).

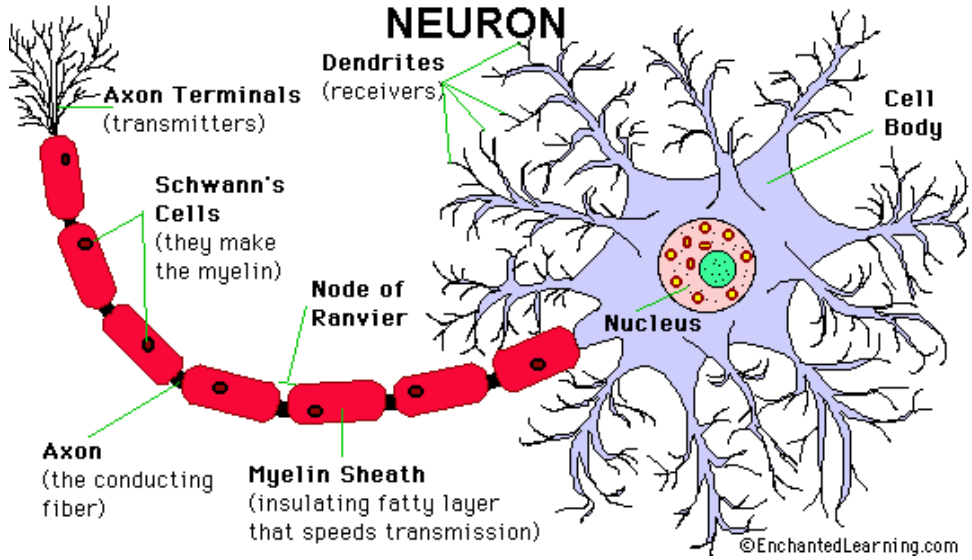


Figure 1-The Anatomy of the Neuron *(Fox, 2016)*

The axon, as opposed to the soma and organelles, is unique to the neuron – no other cells have axons. Axons are characterized by a lack of ribosomes and rough endoplasmic reticulum (rough membrane which holds ribosomes) and a different protein composition than the soma. Ribosomes are crucial to the functioning of the cell, as they synthesize new proteins from amino acids and information transferred by messenger RNA. Because no protein synthesis occurs in the axon due to a lack of ribosomes, all of the proteins found in the axon are made in the soma. The different proteins found in the membrane of the axon allow it to transfer information over great distances. The diameter of the axon is very important, as the thicker the axonal diameter, the faster the impulse travels down the axon.

The axon is made up of three parts: the axon hillock, which is the start of the axon; the axon proper, which is the middle portion of the axon, and the axon terminal, the end of the axon. The terminal is the site where the axon comes into contact with other neurons at a point called the synapse. When a neuron comes into synaptic contact with another cell, it is said to have innervated the other cell. The cytoplasm of the axon terminal is distinct from that of the axon in that a) the cytoplasm in the axon terminal does not contain microtubules; b) the terminal contains bubbles of membrane, called synaptic vesicles; c) the inside surface

that faces the synapse of the membrane contains far more proteins than normal, and d) it has many mitochondria (the powerhouse of the cell), suggesting a high energy demand (Bear).

The synapse itself is divided into two sides: the presynaptic side, and the postsynaptic side. Synaptic transmission, or the transmission of information via a synapse, works by using electrical and chemical signals. The presynaptic side is often the axon terminal of a neuron, while the postsynaptic side is usually the dendrite or soma of another neuron. First, an electrical signal travels through the axon to the axon terminal. When it arrives, the electrical signal is converted into chemicals, which then cross the synaptic cleft (the space between the two sides of the synapse). Once on the postsynaptic side, the chemicals are then coded back into an electrical signal, where they then continue until they meet another synapse, and the process repeats. The chemical substance that crosses the synaptic cleft is known as the neurotransmitter and differs in different types of neurons. Many toxins and drugs work in the synapse, impairing the transmission of sensory information (Bear).

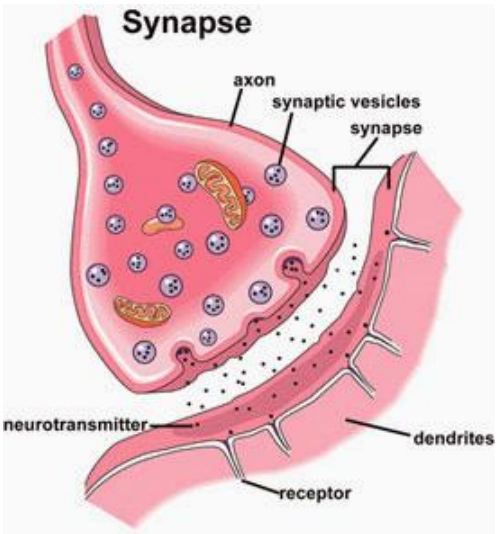


Figure 2-The Anatomy of the Synapse *(Medical Terms, 2016)*

The dendrites of a neuron are often collectively called the dendritic tree. Because the dendrite is often the postsynaptic side of the information transmission, dendrites are covered by synapses. The dendritic (postsynaptic) membrane has many specialized protein cells known as receptors that detect neurotransmitters from the synaptic cleft. Glia, which are currently suspected to have a major role in the process of information transmission in the brain, perform a very important role in the form of oligodendroglial and Schwann cells. These two types of glial cells create layers of membrane around axons, using a material called myelin. The importance of the myelin sheath around axons is that it significantly increases the speed of neuronal transmission, as it speeds the propagation of nerve impulses.

The most important aspect of nerve transmission is the action potential, which allows the sensory information to be transmitted at the synapse to another synapse. The way an action potential works involves many different parts of the neuron and relies on several background settings. It is important to talk about the three main actors in the action potential: the salty fluids on both sides of the membrane, the membrane itself, and the proteins that

span the membrane. First, because water is a polar covalent molecule (meaning that different parts of the water molecule have different charges), and because water makes up most of the cytosol, other polar molecules tend to easily dissolve in the cytosol. An ion is an atom or molecule with a net charge; a cation has a net positive charge, while an anion has a net negative charge. The most important ions in the action potential are sodium (+), potassium (+), calcium (+), and chloride (-) ions (Bear).

Substances that dissolve in water because of their uneven electrical charges are said to love water, or be hydrophilic. Substances that are bonded covalently that are nonpolar are said to be hydrophobic, as they do not dissolve within water (such as lipids). The main chemical building blocks of the cell membrane are phospholipids. Phospholipids are made up of nonpolar chains of carbon and hydrogen atoms, and polar phosphate groups. The polar "head" (the phosphate group) and the nonpolar "tail" (the carbon and hydrogen) combine with another layer of the same phospholipids to create the phospholipid bilayer. In this layer, the hydrophilic heads point outward, one side into the cell and the other out of the cell, and the hydrophobic tails face each other. This stable arrangement serves to isolate the cytosol within the cell from the extracellular fluid outside the cell.

The importance of proteins cannot be overstated, as they are used to create channels in the phospholipid bilayer to allow fluids to pass through. A functional ion channel across the membrane requires around 4-6 proteins to bind together with a pore in between them. These ion channels allow for the flow of ions between the cytosol and the extracellular fluid. There are many important characteristics of ion channels, including ion selectivity-many ion channels can only transfer one type of ion, such as potassium or sodium, and gating, which allows the channel to be opened and closed at different times. Another way that proteins can transport ions across the phospholipid bilayer is through the use of ion pumps. An ion pump utilizes membrane spanning proteins and the energy left over from the breakdown of adenosine triphosphate (ATP) to pump ions across the bilayer. These pumps play an important role in neuronal signaling, as they transport sodium and calcium from inside the membrane to the outside. Pumps work against the concentration gradient, because ions would normally flow to equilibrium (diffusion and electrical charge wise). All of the movement across the membrane is possible because the membrane is semipermeable, meaning some substances can go through it, while others cannot (Bear).

There are two different forces that push ions across the membrane – diffusion and electricity. Because the existence of an open channel does not mean that ions must use it, diffusion is used to move ions across the membrane. Diffusion occurs when there is a concentration gradient, meaning there is more of an ion on one side than the other. Diffusion means that the ions will cross the channel to create an equal number of ion on both sides, thereby lowering the concentration gradient (ions move from high concentrations to low concentrations). Electricity is another way to get ions to move, but it does not use a concentration gradient. Because opposites attract with charge, and because ions are charged particles, when an electrical field is created, the cations will rush to the negative charge, and the anions will rush to the positive charge. Both the electrical potential (the force exerted on a charged particle, or the difference in charge between the anode and the cathode) and electrical conductance (the relative ability of an electrical charge to migrate from one point to another) affect the amount of current that will flow. So overall, the movement of ions across the membrane depends on the concentration gradient and the electrical potential across the membrane.

At rest (when the neuron is not firing an action potential), the potential of a neuronal membrane is -65 millivolts (mV). To create this negative resting membrane potential, it is important to look at the role of potassium, a positive ion, and any anion. In

the membrane, imagine an equal amount of K+ and A- (any), and imagine that outside the cell, there is nothing. Initially, K+ will flow out of the cell, because of the concentration gradient and the need to diffuse. A-, as it does not have an ion channel, cannot flow across the membrane. Then, at some point, the charge in the membrane will be so negative that because of the electrical potential, K+ will start coming back into the cell. At some point, the forces of diffusion and electrical potential will be equal and opposite, leaving a state of equilibrium in which the inside of the cell is left negatively charged. This example can be used to make four major points: 1) a miniscule change in ionic concentration (the K+ leaving the cell) can cause large changes in membrane potential; 2) the net difference in electrical charge is concentrated around the membrane, as both positive and negative can interact through the bilayer (because it is so narrow) and so are pulled to each other; 3) ions are driven across the membrane at a rate proportional to the difference between the membrane potential and equilibrium potential; 4) if the concentration difference across the membrane is known, then the equilibrium potential can be calculated. Every ion has its own equilibrium potential; potassium has an equilibrium potential close to -80 mV, which is close to the resting potential of the cell, but sodium plays a role, lowering the equilibrium potential of the cell as a whole to -65 mV (Bear).

One of the most important pumps in regard to the ionic concentration gradient is the sodium potassium pump, which uses the energy of ATP to ensure that there is more potassium inside the cell than outside the cell, and that there is more sodium outside the cell than inside. These ion pumps restore the original concentration levels of ten times more sodium outside the cell than inside, and twenty times more potassium inside than outside. However, differing from the previous example, the cell is permeable to several different ions, mainly K+ and Na+. Because of this, one would assume the equilibrium potential to be somewhat of an average between that of K+ and Na+. However, the cell is about fourty times more permeable to potassium ions than sodium ions, and so the equilibrium potential of a cell at rest tends to be around -65 mV.

The action potential is made up of three distinct parts. First, the rising phase, in which there is a quick depolarization of the membrane, from around -65 mV to 40 mV. When the inside of the neuron is positively charged compared to the outside of the cell, it is called the overshoot. The falling phase of the action potential is when the cell quickly repolarizes, often characterized by becoming even more negative than the resting potential. Gradually, the cell depolarizes until it is at resting potential. Every action potential takes a total of around 2 milliseconds.

The action potential can only be achieved, however, if the cell depolarizes to a certain benchmark. When a stimulus is felt, and the cell depolarizes, an action potential will only be triggered if the generator potential (depolarization of the cell) reaches a certain level of depolarization – the threshold. Action potentials can be compared to taking a picture on a camera – you can push and push, but if you don't push the shutter button past a certain point, no picture will be taken; however, once it reaches that point, the *effect is the same* – the picture is not "longer" nor is it different than any other pressure above the threshold. This is effectively saying that the feeling of pain is not determined by the size of the action potential, as all action potentials are the exact same. The feeling of pain is determined by which fibers are being activated and how many action potentials are fired.

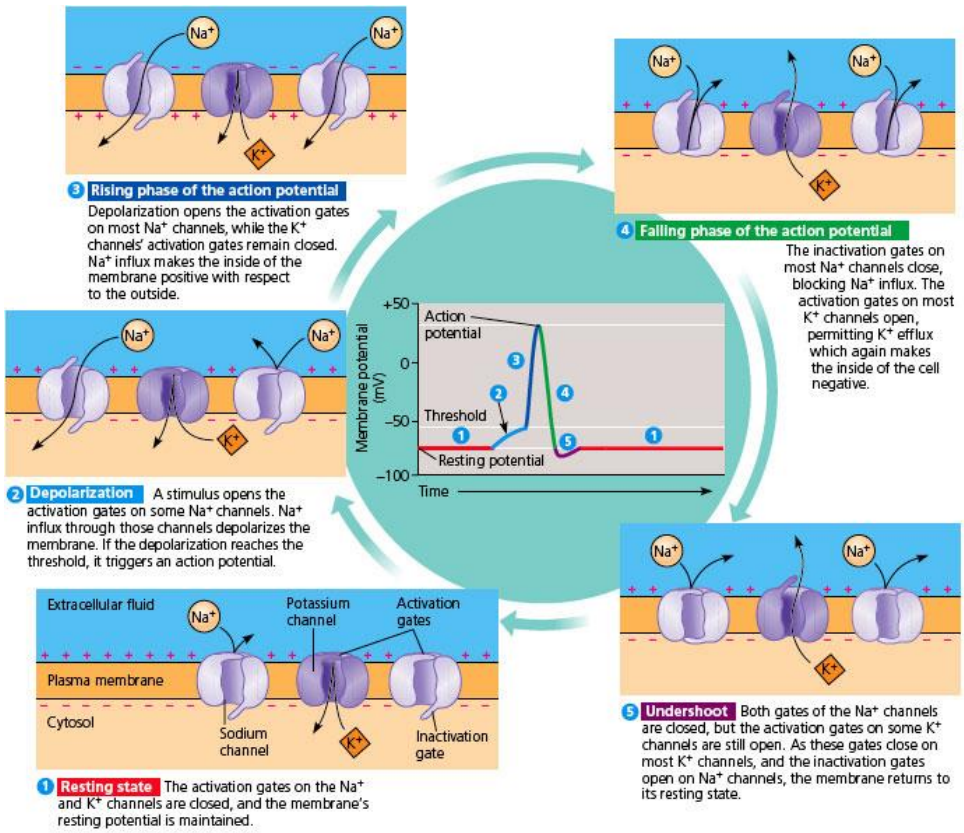


Figure 3-The Sequence of the Action Potential *(Notes on Medicine/Surgery, 2012)*

PAIN PERCEPTION AND INHIBITION

There are two main nerve fibers responsible for the transmission of pain information, the A delta and C fibers. There are several differences between the pain fibers and other sensation fibers as well as between the two pain fibers. The most important difference is that touch fibers have specialized structures (such as Pacinian corpuscles) at their ends, while pain fibers have free nerve endings, which form complex networks that are the sensory receptors for pain, called nociceptors. Nociceptors only trigger when a sensation is enough to cause harm to the body (an injury). The biggest differences between the A delta fibers and the C fibers are the diameter of each and the presence or absence of the myelin sheath. The A delta fiber, referred to as fast pain, is characterized by a wide diameter, a thick myelin sheath, very fast pain transmission times, and acute pain. On the other hand, C fibers are referred to as slow pain, characterized by a narrow diameter, no myelin sheath, slow pain transmission times, and chronic pain. Because of the difference in speed transmission, oftentimes when people are injured, they feel a sharp pain very quickly, and then a dull ache takes its place, often lasting much longer than the sharp pain. This happens because of the lag time between the brain receiving information from A delta fibers (fast pain), and the brain receiving information from C fibers (slow pain). One of the most interesting theories related to the difference in transmission of pain between fast and slow fibers is the Gate Control Theory, put forward by Melzack and Wall.

Type of Nerve Fibre	Information Carried	Myelin Sheath?	Diameter (micrometers)	Conduction Speed (m/s)
A-alpha	proprioception	myelinated	13 - 20	80 - 120 
A-beta	touch	myelinated	6 - 12	35 – 90 
A-delta	pain (mechanical and thermal)	myelinated	1 - 5	5 – 40 
C	pain (mechanical, thermal, and chemical)	non-myelinated	0.2 - 1.5	0.5 - 2 

Figure 4-Differences in Touch and Pain Fibers *(Dubuc, 2016)*

The Gate Control Theory of Pain relates to the existence of T cells in the spinal cord, which are located along the pathway that pain fibers use to transmit pain sensation to the brain. The T cells in the spinal cord send pain messages to the brain, as they are excited by both large and small diameter nerve fibers. However, along the way, inhibitory neurons in the substantia gelatinosa (SG) inhibit the T cells, reducing the pain transmission to the brain. The basic idea of the gate control theory is that because there are inhibitory cells that act as gates (blocking the pain if closed, allowing it to continue if open), the fast sensation will "beat out" the slow sensation. Because the fastest fibers are touch fibers (they are the largest in diameter, and by far the most myelinated), the touch sensation will "override" the pain sensation. This theory provides a physiological base to the study of pain and also helps to explain natural phenomena: why does rubbing a body part after hitting it lessen the pain? It is because there are gates, and the faster nerve fibers win out in the race to the brain because of the inhibitory cells that block pain when touch is activated. This theory also opens the door to the idea that pain in itself is affected by the brain and its reactions to stimuli and encourages the study of one's own culture and its influence on pain perception.

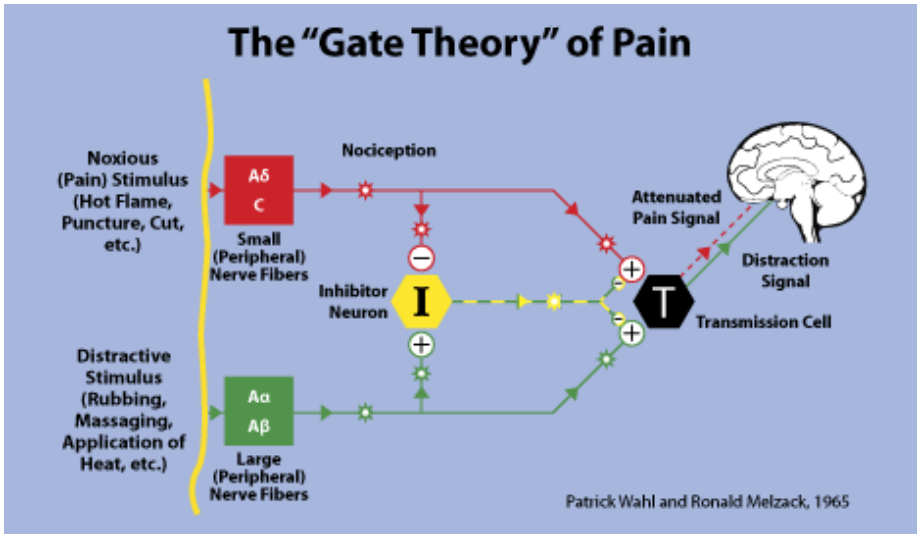


Figure 5-The Gate Control Theory of Pain *(Genesis Medical Clinic, 2014)*

Another important influence on pain perception are opioids. Opioids are used to relieve pain, and they do this by blocking receptor sites for pain. Opioids, and opiates, help to relieve the pain by blocking nociceptive reception sites, thereby blocking the pain signals from the brain. Some, called endorphins, are naturally produced and are used to prolong survival. Endorphins are extremely helpful in nature, as humans evolved to have endorphins at a time when a serious pain in the wild might signal death. For example, when a human was running from an animal and tripped, fell, and broke his ankle, the pain would normally make him lie dormant until the animal reached him and attacked. The brain, in its yet to be understood wisdom, actively releases endorphins to block the pain from reaching it, letting the man run on until he is safe, at which time the pain receptor sites will be unblocked, and he will feel the full pain of his injury.

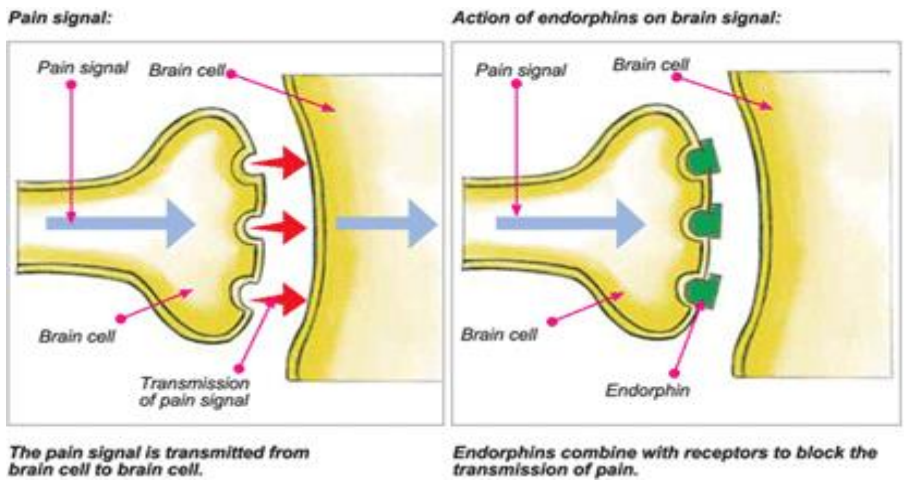


Figure 6-Endorphins Blocking Pain Transmission *(Martinez, 2012)*

Opioids can block pain in two different manners: presynaptic inhibition and postsynaptic inhibition. Postsynaptic inhibition occurs when an endorphin is coded to release K+ in a cell, causing an extreme hyperpolarization of the cell, making it almost impossible for an action potential to occur. Presynaptic (axoaxonic) inhibition occurs when an opioid neuron closes Ca2+ channels, prohibiting the release of a neurotransmitter substance, and inhibiting the feeling of pain. Endorphins are naturally released through pain, injury, physical stress, food, and pleasure. Because the use of opioids to block pain also activates the pleasure pathways, opioids are extremely addictive. They provide users with a sense of euphoria, and, according to the National Alliance of Advocates for Buprenorphine Treatment, "Opioids target the brain's reward system by flooding the circuit with dopamine. The overstimulation of this system, which rewards our natural behaviors, produces the euphoric effects sought by people who misuse drugs and teaches them to repeat the behavior." The use of opioids to block pain activates receptor sites because they mimic the structure of a normal neurotransmitter, which causes a strong feeling of euphoria, leading to addiction and a need to increase dosage of the opioid with every usage.

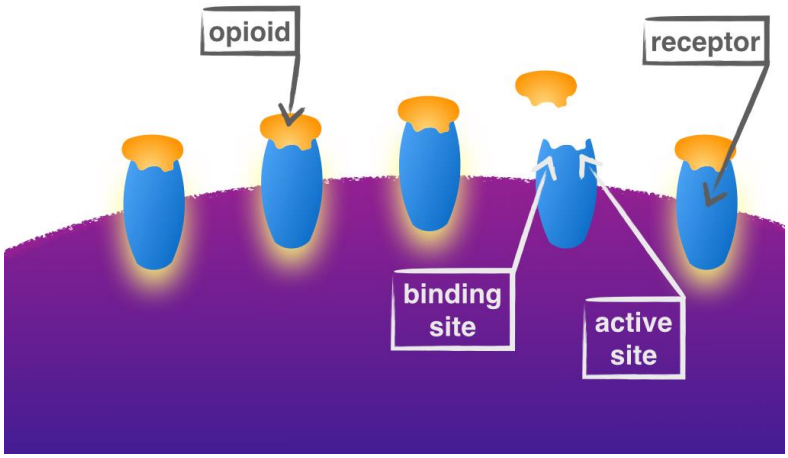


Figure 7-Opioid Reception in the Brain *(Information for EMS, 2016)*

Naloxone, a drug that is used in emergency overdose situations of opioids, effectively blocks the opioids by binding to and blocking receptor sites. When naloxone is administered to a patient, it travels through the same pathways used by endorphins and opioids and binds to the same receptor sites. However, instead of producing a similar reaction of calming and pain reduction, naloxone merely blocks these reactions, causing the full force of pain to be felt and also saving the life of someone overdosing from an opiate. Naloxone, also known as Narcan®, is used very widely because it has no potential for addiction, and has no negative effects in the body even if opioids are not present in receptor sites in the brain (Harm Reduction, 2016). Naloxone blocks the opioids because it has a stronger affinity to the receptor sites in the brain, which effectively displaces the opioid molecules from the receptor sites, negating the opioid effect.

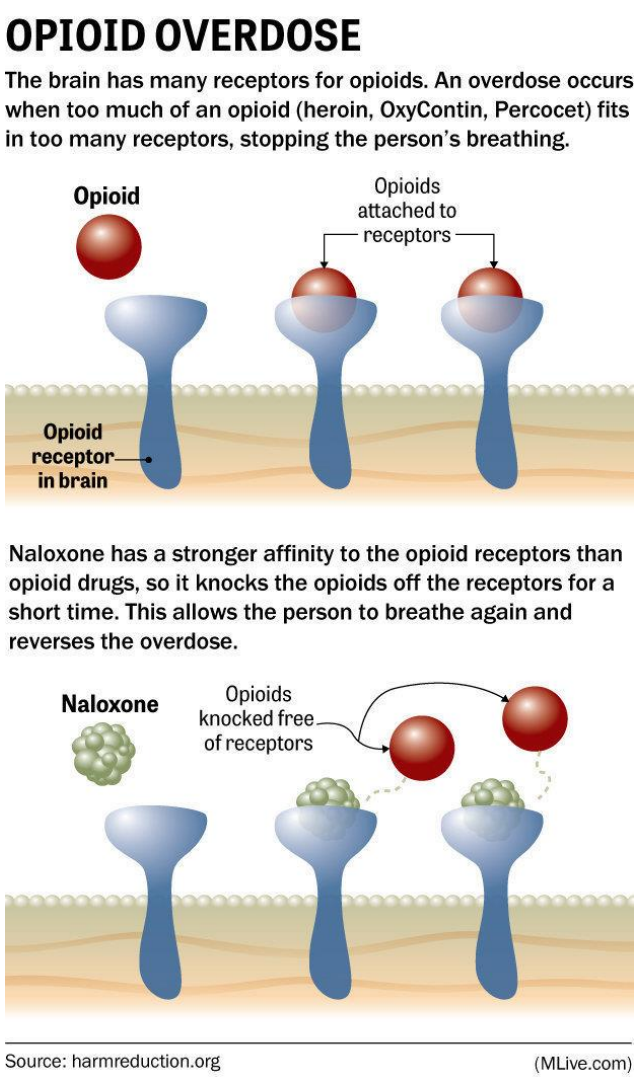


Figure 8-Naloxone and Opioid Overdose *(Thoms, 2015)*

The placebo effect is another fascinating area of pain inhibition and involves endorphins produced by the body. The placebo effect is the change in a person's symptoms when receiving a placebo (American Cancer Society, 2015). A placebo is an inactive substance that is often branded as a medication in order to improve patient symptoms without using actual medication. Often, placebos are pills made of sugar or syrups made of water. Placebos are designed to seem like a real treatment, but one that doesn't directly affect the illness (American Cancer Society, 2015).

The most important breakthrough in placebo research came in a study done in 1978 and published in The Lancet journal entitled "The Mechanism of Placebo Analgesia." In this study, dentists analyzed the effects that placebos had on people's perceptions of pain following dental surgery. Hours after the surgery, there were two rounds of medicating.

> 3 hours and 4 hours after surgery naloxone or a placebo was given under randomised, double-blind conditions. Pain was evaluated on a visual analogue scale. Patients given naloxone reported significantly greater pain than those given placebo. Patients given placebo as their first drug were either placebo responders, whose pain was reduced or unchanged, or nonresponders whose pain increased. Naloxone given as a second drug produced no additional increase in pain levels in nonresponders but did increase pain levels of placebo responders. Nonresponders had a final mean pain rating identical to that of responders who received naloxone as their second drug. Thus the enhancement of reported pain produced by naloxone can be entirely accounted for by its effect on placebo responders. These data are consistent with the hypothesis that endorphin release mediates placebo analgesia for dental postoperative pain. (Levine, 1978.)

The study found that the effects of the placebo in responders (meaning the placebo initially reduced their levels of pain) were completely removed following administration of naloxone, an opioid antagonist. This means that the placebo effect, specifically proven by this study in postoperative dental pain, can be completely negated when opioids are blocked. This sheds light on the problem of the placebo effect and how it works, proving that the placebo effect is determined by the release of endorphins produced by the brain, which then bind to opioid receptor sites, blocking the sensation of pain. When naloxone is administered, however, the endorphins cannot reach the receptor sites, causing an increase in pain sensation (Feinberg, 2013).

MEASURING PAIN

There are many different scaling methods used in studies on pain perception, so it is important to standardize the way that pain is measured. The problem with pain in the laboratory is that there are many ways to measure it and not a lot of ways to check to make sure that the reading is accurate. Many studies take into account how people say they feel when measuring a noxious stimulus. However, does the exposure to a noxious stimulus prior to the experiment affect the actual pain felt by the subject? If a study is testing pain by using electrical shocks, and one of the participants is an electrician and so is used to the shocks, will he report feeling less pain, even if the pain has the same physiological effect on him as on everyone else in the study? The first problem with measuring pain is that in studies measuring pain, there is a "persistent failure of subjects to relate stimulus intensity to pain intensity… [This] is one of the strongest reasons to question the classical attempt to group pain with the familiar sensations evoked by external sources" (Wall, 1979).

One way to measure pain is by taking a cognitive approach. Melzack and Torgerson (1971) expanded the ability to measure pain by evaluating the clinical descriptions of pain. They chose 102 words to describe pain, breaking the adjectives into three categories and thirteen subcategories. From this list, the words were scaled by how much pain each word represents, and then based on this analysis, Melzack created the McGill Pain Questionnaire, which is used to describe the multiple components of the pain experience. Unfortunately, a cognitive approach to measuring pain is problematic since culture and other factors can affect cognition itself and therefore the perception of pain (Turner and Chapman, 1982, Melzack and Wall, 1983, Fouberg, 2009).

In addition to using a cognitive approach, other researchers attempted to measure pain through the use of numerical scales. One important scale used in the distinction of pain is the Borg perceived pain scale, or the CR (category-ratio) 10 scale. The Borg perceived pain scale (similar in nature to the Borg perceived exertion scale) is simply a 0-10 scale to

say how much pain is hurting. Such scales present certain problems: first, pain perception in a number is built from a great number of different factors, many of which will be discussed in this paper; second, a scale that is finite reduces the sensitivity of the number before and after intervention, and so is often unreliable (Lundeberg 2001).

1-10 Borg Scale of Perceived Exertion	
0	Rest
1	Really Easy
2	Easy
3	Moderate
4	Sort of Hard
5	Hard
6	
7	Really Hard
8	
9	Really, Really Hard
10	Maximal

Figure 9-Borg Perceived Exertion (and Pain) Scale *(Woodruff 2013)*

Because of this, researchers in Sweden focused on testing the Reliability and Responsiveness of Three Different Pain Assessments, including the visual analogue scale (VAS, 0-100 scale), the numerical rating scale (NRS, 0-20), and a new type of pain scale called the magnitude matching scale. The researchers understood that having a scale with numbers and ends is not a good way to measure pain, since countless factors go into a single number to describe pain – was the person raised to talk about pain or hide it? Was the person raised to believe that every little injury was dangerous or was he less fearful? Was the culture the person was raised in determined to investigate the true feelings of pain or to hide them? Is the personality of a person such that it changes how he would answer the question? Because of these questions and many more, researchers in Sweden decided to test pain by using magnitude matching, or painmatching. Simply put, magnitude matching uses an electrode box to slowly increase the pain of the shock from the electrode box and then stop the shocks when the pain of the shocks is equal to the pain of the injury. The painmatcher technique showed a greater reliability in the true determination of pain by creating a more reliable and stable way of measuring pain and comparing pain after treatment with pain before treatment.

Other scales are used in clinical settings. Five main scales are used for the description of pain by nurses, a group that is required daily to evaluate the extent of a patient's pain, and determine if it is tolerable or not. The first of these five is the FLACC scale (Face, Legs, Activity, Cry, Consolability). This scale is used mainly for children and is used by the medical community to determine the pain by assigning two points (and characteristics) to each category. The Visual Analogue Scale (VAS, as discussed above) is a straight line on which patients are supposed to draw an X to mark the location on the line for the level of pain they are feeling. The FACES scale is merely a series of cartoons

depicting people in pain, and the subject is supposed to point out the level of pain based on the expression of the face. The last scale is the Numeric Pain Intensity Scale, which is simply a 0-10 scale, in which the subject is asked how extreme his pain is.

In spite of the careful efforts to create and test reliable instruments with which to measure pain, these efforts will be limited due to the effect that experience and particularly culture can have on the perception of pain. That issue will be taken up in the next section.

THE EFFECT OF CULTURE ON PAIN PERCEPTION

Many studies have shown that both culture and the diversion of thoughts can lessen the experience of pain. In a variety of different studies, it has been shown that diverting attention away from noxious stimuli lessens pain, but mainly benefits those with already high pain thresholds, leaving those that experience the most pain with the least benefit from diversion of thoughts (Anderson 2002, Anderson 2009, Campbell 2012, Juarez 1998, Juarez 1999). One possible explanation for this is that the subjects with higher pain thresholds were already using self-taught techniques to manage the pain. This could mean that a given culture teaches certain techniques in order to lessen pain or to solve the cause of the pain.

One of the most important observations on pain relating to situation and culture, as written by Turner and Chapman (1982), is that "the cognitions (attitudes, beliefs, and expectations) people maintain in certain situations can determine their emotional and behavioral reactions to those situations. As cognitive (e.g., distraction, significance of the pain for the individual) and emotional variables (e.g., anxiety) influence the experience of pain, it seems logical that the modification of cognitions could be used to alter the pain experience." As culture itself is defined as a group of belief systems, norms, and values practiced by people (Fouberg, 2009), the ability of culture to change the perception of pain has been documented for decades (Melzack and Wall, 1983).

Thus, the problem with scales based solely on numbers and faces – really all mentioned scales except for the painmatching method – is that many cultures have different experiences with pain and differences in how they use and think of numbers, and so the numbers are not always reliable. In regions of the world where a common response to pain is stoicism, the lack of expression of the pain often relates to underreporting the pain and can also cause true pain to go undiagnosed. The faces provide a similar problem – in some Asian cultures, smiling is not a form of happiness; it often is used to show anger or embarrassment (Carteret, 2011). An additional issue when using descriptive words to describe pain is that many cultures have completely different words to describe the pain than others. For example, in the United States, a doctor may ask if the pain is sharp or dull. In another culture, a person may describe pain in terms of natural symbols, such as like a bolt of lightning, or a spider web (Burhansstipanov, 2000). In addition, different cultures perceive the approach to dealing with illnesses differently. In a study of Native Alaskans living on U.S. reservations, Burhansstipanov described the various barriers to dealing with illnesses becoming prevalent in the native community. Cancer specifically was referred to by Native Americans as a "white man's" disease. A diagnosis of cancer in a native community results in depression and also a feeling of closure – "the cancer will take its natural path, whether or not I will heal."

This is a common story among the thousands of cultures in the world and has an effect on how these cultures perceive of and treat pain. Many do not want to give up their cultural practices and traditions of remedying pain in favor of the Western method of pain treatment. Especially in tribal cultures, native healers wish to maintain cultural practices and avoid the "impersonality" of Western medicine (Pascale, 2008). Every culture has developed methods of pain management and created ways to tolerate and express pain.

Several to be discussed in this paper have become prevalent in mainstream society. For the purpose of this research, only management techniques dealing with chronic pain will be discussed.

MANAGING PAIN

Acupuncture is one of the most interesting methods of pain management, as it has become extremely important in both folk and popular medicine, while being used for different reasons. Acupuncture, as developed by the Chinese, is believed to balance the flow of life force in the body, called chi (Mayo Clinic, 2015). In contrast, Western medicine believes acupuncture stimulates nerves and connective tissue, increasing blood flow. Because acupuncture is believed to have positive health effects, Western medicine has incorporated it – at least partially – in what is now called medical acupuncture: licensed Western doctors who are also trained to practice acupuncture (Rotchford, 2016).

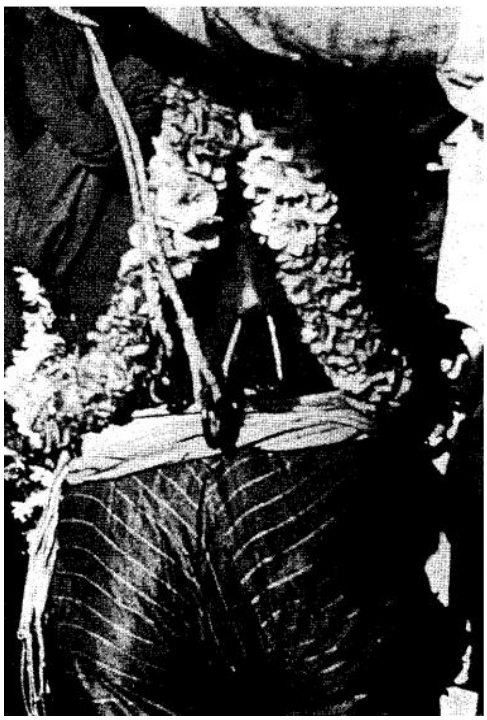
However, many pain management techniques that are widespread in certain cultures have not reached Western medicine or are regarded as illegitimate practices that have no actual impact on pain and work only as placebos. But, perhaps more interesting than consciously developed methods of reducing pain are the subconscious ways in which a culture shapes people's physiology to the point that they respond to pain in different manners than others would. Take, for instance, the Satere-Mawe tribe, a group native to the Amazon rainforest. In preparation for adulthood, all boys must wear gloves made of leaves, in which bullet ants, the most painful in the world, are woven into the linings (Cultures at Customs, 2013). Each boy must wear the gloves for a total of ten minutes at least twenty times.

To any who are not of the Satere-Mawe tribe, the pain of a single bite from the bullet ant is enough to last for several hours, even days. The bullet ant contains a powerful toxin that causes shaking, interrupts the nervous system, and has been called the most painful insect sting in the world by the British Broadcasting Corporation (Zoe Gough, 2015). When an initiate wears the gloves, he is not allowed to show pain – he must not scream, or cry, or show any emotion. However, those outside of the culture can barely tolerate the pain. As soon as an Australian man tried to put the gloves on, the pain overcame him. Within ten seconds, the gloves were back off, and the pain was so severe that he was taken to the hospital. Even after an entire day had passed, he could barely move his fingers. What, then, allows the natives to tolerate the bullet ants, while non-natives cannot last more than a few seconds? As explained by Carl Cotman in "Behavioral Neuroscience: An Introduction", the perception and meaning of pain varies within cultures.

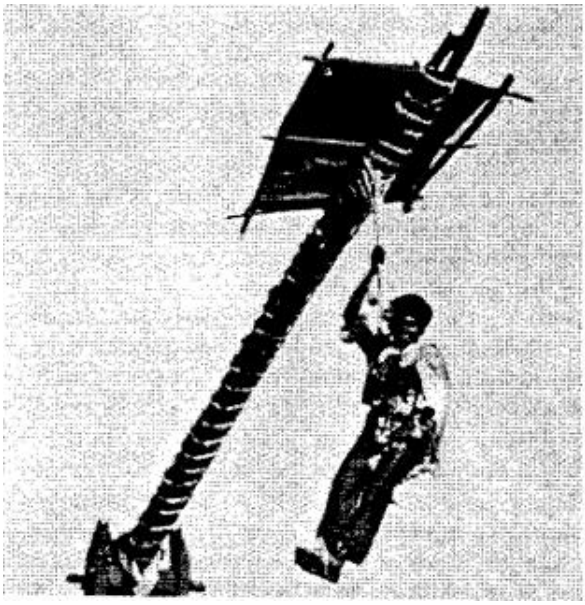
Picture 1-Satere-Mawe Initiate Wears the Bullet Ant Gloves *(Cultures and Customs 2013)*

Picture 2-A Member of the Satere-Mawe Tribe Shows the Inside of the Famed Bullet Ant Gloves During the Initiation Ceremony *(Cultures and Customs)*

Similar to the initiation ceremony of the Satere-Mawe, in several remote villages in India in the Deccan region, metal hooks are hung from posts (D.D. Kosambi, 1967). An ancient ceremony involving a sacred god takes place with these hooks. A man from each village is selected every year to be hung by two hooks from his lower back and left there to swing by his skin. Many outsiders regard this as savage and brutal as well as evidence of the lower status of indigenous peoples of different color. However, to be selected for this honor is not considered savagery by those who practice it and is in fact one of the greatest honors one can receive. In fact, the men chosen for this honor report *no pain whatsoever*, even though a metal hook is being dug through the skin in the back, and the men are being left to hang (Kosambi, 1967). Even following the ceremony, the selected receive no special medical treatment. The wounds are merely treated with wood ash during the ceremony, and within two weeks, the wounds are reported to have completely closed and have ceased to be visible. To anyone not familiar with the honor of the hook hanging, being hung by hooks would be incredibly painful, but in these select villages in rural India, there is no pain reported.



Picture 3-An Indian Man About to Be Hung from A Pole by Rope Attached to the Hooks in His Back *(Kosambi, 1967)*



Picture 4-An Indian Man Hangs from a Rope Attached to Metal Hooks in His Back
(Chattopadhyaya, 1967)

These fascinating cultural experiences show the amazing effects that culture and nurture have on the physiology of a person. Because of the honors associated with the traditions of the Satere-Mawe and the Indian villages, the participants feel little pain, and it does not last. Compared to the effects that these traditions have on foreigners, it is miraculous to think that culture can have such an astounding impact on the way one feels pain. Thus, to determine the effect that culture has on pain perception, it is important to emphasize the placebo effect, endorphins, and the role of naloxone.

UNDERSTANDING THE ROLE OF CULTURE IN PAIN PERCEPTION

How does culture play a role in managing pain? What is the physiological mechanism that a culture can trigger to allow native people to tolerate the extreme pain caused by their cultural practices? In order to determine the effect that culture has on the perception of pain by these two cultural groups, an experiment must be conducted in which culture is isolated as the one variable in the study. One way to accomplish this would be through the use of naloxone. Considering that naloxone blocks the effects of endorphins (and the placebo effect), it is possible to isolate cultural influences by administering naloxone prior to these cultural practices. In order to determine the effect that culture has in blocking the feelings of pain, naloxone must be used to stop the pain blocking effect of endorphins naturally produced by the brain. Exposure to naloxone would then block any endorphins produced by the brain and allow the study of whether or not the pain is then felt. If the pain is then felt by members of these cultural groups, it becomes apparent that their cultures cause them to produce endorphins to block the feelings of pain because of the significance of the customs. If, on the other hand, they do not feel the pain, it then becomes necessary to evaluate what stops members of these cultural groups from feeling pain, since it is not the production of endorphins that blocks pain.

In addition to these two experimental groups, it is necessary to include control groups of those born outside the culture, who would thus not be subject to the same cultural influences on pain perception. The introduction of control groups made up of outsiders would allow researchers to study how culture influences endorphin production. If all the subjects in the control group and the experimental groups do not feel the pain, it can be ascertained that culture is *not* the factor blocking the pain. If the subjects in the control group do feel the pain, while the subjects in the experimental group, which have not received naloxone, feel no pain, it can be ascertained that endorphins are produced by the brain because of the culture one is born into. This would prove that culture has deep impacts on the physiology of a person. If members of the Satere-Mawe and Indian tribal groups are given naloxone during the rituals detailed above, they will feel the pain, since the endorphins that normally block pain will be blocked.

This experiment will take the practices of two cultures on different sides of the world in order to isolate culture as a single variable and its impacts on pain perception. Because the effects of naloxone and endorphins are well known (at least compared with knowledge about the influence of culture on pain), by isolating culture as the only variable in the experiment as a whole, the true impact of culture will be able to be seen on the physiology of a person. With this knowledge, scientists around the world could begin transforming the way that we learn about pain and delve deeper into aspects of human physiology previously unknown.

METHODS

In order to ensure accuracy of results, it is necessary to have four different groups for each culture, so eight groups in total. The first group will be the control group, made up of outsiders who are not members of the culture in which the initiation ceremonies are practiced. The second group will again be made up of outsiders, but will instead be administered naloxone, in order to determine if all who experience these cultural practices do not feel the pain. Next, the third group will contain members of the cultural group, but they will not be administered naloxone before, during, or after the ceremonies. Lastly, the fourth group, the main experimental group, will be made up of members of the cultural group, and they will be administered naloxone before, during, and after the initiation ceremonies.

When choosing a scaling method for the measurement of pain, the painmatching technique allows the greatest reliability of results, as it removes many biases. The procedure will be as follows.

For the control group, there will be no need for testing prior to the initiation, since the only pain measurement needed will come during and after the occurrence of pain (while the gloves are on the hands of the subject, and while the hook is in the back, and after both of those events). During the initiation ceremony, subjects will be administered electric shocks, increasing in value until they match the pain felt during the ceremony (these electric shocks will be administered in the same way as detailed by the lab report of The Reliability and Responsiveness of Three Different Pain Assessments). The control group wearing gloves made of bullet ants will have the pain measured twice during the ceremony, once after one minute, and once again after eight minutes. The control group with the metal hook, however, due to the increased time spent during the ceremony, will have pain measured four times during the ceremony, once on the first day, again on the fifth day, again on the ninth day, and lastly on the fourteenth day. Following the procedure, at one hour after, one day, two days, three days, seven days, and fourteen days, the pain again will be measured in this same method, in order to evaluate the longevity of the pain in the subjects, as well as the

decrease of pain over a period of two weeks. This will then be juxtaposed with the longevity, intensity, and change in pain sensation to the other two groups.

The second group will follow the same pain measurement timelines as the control group and will also be made up of foreigners to the cultures being studied. However, the second group will be administered naloxone in order to block the possible effects of endorphins being released by the brains of non-culture members. The second group will be administered naloxone on different timelines depending on the ceremony. For the Satere-Mawe initiation, the naloxone will be administered immediately prior to putting on the gloves and immediately following the putting on of the gloves. For the Indian ritual, the naloxone will be administered immediately before the hook is inserted into the back of the subject and will then continue to be administered twice daily for the remainder of the two weeks, and then the next two days following the ceremony.

In order to test if endorphins, and culture, are actually the cause of the lack of pain, it is necessary to have two different groups both made up of members of the culture practicing the initiation ceremonies. The first of these two groups, detailed in this paragraph, will follow the normal procedure for the ceremonies of that culture. They will not be administered naloxone or any other drug, and will simply follow the normal rules of their ceremonies. This group will have pain measured at the same time as those of the other groups, as it is necessary to be able to compare results to have each group have pain measurements from the same times.

The final group, made up of members of the culture, will be administered naloxone with the same schedule as the second group, and will also have pain measured at the same times. These four groups allow for each variable to be isolated, and thus to determine if it is the cause of the lack of pain felt by members of these two different cultural practices.

Lastly, it is important that these practices take place in the usual locations – the studies will not be conducted in the laboratory, meaning that the equipment needed must be transported to both Amazonia and the Dhakan region of India. If the experiment takes place outside of the cultural context in which it normally occurs, the results of the experiment are jeopardized, because the lack of this cultural context may take away the cultural significance of the event, preventing any release of endorphins by the native group that is not being administered naloxone, making the results not effective.

	Group 1	Group 2	Group 3	Group 4
Cultural			X	X
Foreigner	X	X		
Naloxone		X		X
No naloxone	X		X	
Naloxone administered (Times)	N/A	Immediately before and after putting on the gloves	N/A	Immediately before and after putting on the gloves
Pain measured (Times)	One and eight minutes into the ceremony; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony	One and eight minutes into the ceremony; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony	One and eight minutes into the ceremony; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony	One and eight minutes into the ceremony; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony

Table 1-Groups and Times for Satere-Mawe Experiment

	Group 1	Group 2	Group 3	Group 4
Cultural			X	X
Foreigner	X	X		
Naloxone		X		X
No naloxone	X		X	
Naloxone administered (Times)	N/A	Immediately before the hook is inserted into the subject, and then twice daily for the rest of the ceremony	N/A	Immediately before the hook is inserted into the subject, and then twice daily for the rest of the ceremony
Pain measured (Times)	The first day of the ceremony, the fifth day, the ninth day, and the 14th day; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony	The first day of the ceremony, the fifth day, the ninth day, and the 14th day; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony	The first day of the ceremony, the fifth day, the ninth day, and the 14th day; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony	The first day of the ceremony, the fifth day, the ninth day, and the 14th day; one hour, one day, two days, three days, seven days, and fourteen days after the ceremony

Table 2-Groups and Times for Indian Experiment

POSSIBLE RESULTS

There are many possible results to gather from the experiment, and so it is important to evaluate all possible scenarios. If none of the groups experiences pain whatsoever in either cultural practice, it can be determined that culture is not the factor prohibiting the feeling of pain and neither are endorphins. If the groups not administered naloxone do not feel pain (Groups 1 and 3), while the groups receiving naloxone do feel pain (Groups 2 and 4), it is highly possible that the blocking of pain regardless of culture occurs because of the presence of endorphins, and that culture is not the factor that reduces pain, since there is a direct correlation between naloxone and perception of pain. If Group 3 experiences no pain, while all three other groups do experience pain, it can be assumed that culture does produce endorphins which serve to block pain, as both of the non-cultural groups felt pain (isolating culture), and the cultural group that received naloxone felt pain as well (isolating naloxone and endorphins). These are the most likely results, as previous

cultural interactions have shown that natives are far less likely, or do not at all, feel the pain of these cultural practices.

In order to fully utilize the data received through this experiment, it is important to also look at the pain measurements taken during and following the experiment. The most important data to evaluate is the difference in pain one person feels compared to the difference in pain others feel (create an average difference for each group, and then compare the groups). The difference in pain felt by each group over time can also shed light on other ways that the brain is affected by culture, such as faster recovery times or even a simple lessening of pain, even if it is still present.

Overall, it is well known that pain is affected by numerous inputs. Race, gender, socioeconomic status, ethnicity, physiology, and culture all help to shape the way that each person feels and responds to pain. It is important to start utilizing these different methods of study in order to determine how we can use the effects our own brains have on nullifying the effect of pain, and how we can better understand the mechanisms of the most complicated organ in our bodies – the brain.

References

"Action Potentials." *Notes on Medicine/Surgery*. (2012, April 6). Retrieved June 28, 2016 from https://dundeemedstudentnotes.wordpress.com/2012/04/06/action-potentials/

Acupuncture. (2015, February 21). Retrieved June 14, 2016, from http://www.mayoclinic.org/tests-procedures/acupuncture/basics/definition/prc-20020778

Anderson, K. O., Richman, S. P., Hurley, J., Palos, G., Valero, V., Mendoza, T. R., . . . Cleeland, C. S. (2002). Cancer pain management among underserved minority outpatients. Cancer, 94(8), 2295-2304. doi:10.1002/cncr.10414

Anderson, K. O., Green, C. R., & Payne, R. (2009, December). Racial and Ethnic Disparities in Pain: Causes and Consequences of Unequal Care. The Journal of Pain, 10(12), 1187-1204. doi:10.1016/j.jpain.2009.10.002

Alvarado, A. J. (2008). Cultural diversity: Pain beliefs and treatment among Mexican-Americans, African-Americans, Chinese-Americans and Japanese-Americans (Unpublished master's thesis). Ypsilanti, Michigan: Eastern Michigan University.

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: Exploring the Brain* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Binder, M. D., Hirokawa, N., Windhorst, U., &Quevedo, J. N. (2009). Presynaptic Inhibition. In *Encyclopedia of Neuroscience* (1st ed., pp. 3266-3270). Berlin: Springer.

Borg, G. (1998). *Borg's Perceived exertion and pain scales*. Champaign, IL: Human Kinetics.

Burhansstipanov, L. (2000). *Native American Health Issues Lessons Learned from Community-based Interventions, and Overview of Native Cancer Projects* [Scholarly project]. In *University of North Carolina at Chapel Hill*. Retrieved June 19, 2016.

Byrne, J. H. (n.d.). Resting Potentials and Action Potentials. Retrieved June 12, 2016, from http://neuroscience.uth.tmc.edu/s1/chapter01.html

Callister, L. C.. (2003). Cultural Influences on Pain Perceptions and Behaviors. Home Health Care Management & Practice, 15(3), 207-11. Web. Retrieved June 19, 2016.

Campbell, C. M., & Edwards, R. R.. (2012). Ethnic Differences in Pain and Pain Management. Pain Management, 2(3), 219-30. *National Center for Biotechnology Information*. Web. 18 June 2016.

Campbell, C. M., Edwards, R. R., &Fillingim, R. B.. (2005). Ethnic Differences in Responses to Multiple Experimental Pain Stimuli. Pain, 113(1), 20-26. Web. Retrieved June 24, 2016.

Carteret, M. (2011). Cultural Aspects of Pain Management. Retrieved June 17, 2016, from http://www.dimensionsofculture.com/2010/11/cultural-aspects-of-pain-management/

Chattopadhyaya, B., & Kosambi, D. D. (n.d.). D.D. Kosambi Combined Methods in Indology and Other Writings. Retrieved June 25, 2016.

Cotman, C. W., &McGaugh, J. L. (1980). *Behavioral Neuroscience: An Introduction*. New York: Academic Press.

Coyne, J. (2014, August 06). The worst pain known to humans: The “bullet ant” gloves of Brazil. Retrieved June 25, 2016, from https://whyevolutionistrue.wordpress.com/2014/08/06/the-worst-pain-known-to-humans-the-bullet-ant-gloves-of-brazil/

Cultures and Customs A look at traditions around the globe. (2013, November 16). Retrieved June 25, 2016, from https://sites.psu.edu/mgeitnerrcl/2013/11/16/satere-mawe-initiation/

Deardorff, W. W. (n.d.). Modern Ideas: The Gate Control Theory of Chronic Pain. Retrieved June 16, 2016, from http://www.spine-health.com/conditions/chronic-pain/modern-ideas-gate-control-theory-chronic-pain

Dubuc, B. (n.d.). Ascending Pain Pathways and Descending Pain-Control Pathways. Retrieved June 14, 2016, from http://thebrain.mcgill.ca/flash/d/d_03/d_03_cl/d_03_cl_dou/d_03_cl_dou.html

Feinberg, C. (2013, January/February). The Placebo Phenomenon. *Harvard Magazine*. Retrieved June 25, 2016, from http://harvardmagazine.com/2013/01/the-placebo-phenomenon

Fouberg, E. H., Murphy, A. B., &Blij, D. (2009). Human Geography: People, Place, and Culture. Hoboken, NJ: J. Wiley.

Fox, S. I. (2016). Nervous System Neurons and Synapses. Our Physiology Group. Retrieved July 18, 2016

Gough, Z. (2015, March 13). The world's most painful insect sting. Retrieved June 25, 2016, from http://www.bbc.com/earth/story/20150312-the-worlds-most-painful-insect-sting

How do opioids affect the brain and body? (2014, November). Retrieved June 19, 2016, from https://www.drugabuse.gov/publications/research-reports/prescription-drugs/opioids/how-do-opioids-affect-brain-body

How do opioids work in the brain? (2008, December). Retrieved June 19, 2016, from http://www.naabt.org/faq_answers.cfm?ID=6

"Information for EMS." (n.d.) *Naloxone and Overdose Prevention Education Program of Rhode Island*. Web. Retrieved June 28, 2016.

Juarez, G., Ferrell, B., &Borneman, T. (1998, September). Influence of Culture on Cancer Pain Management in Hispanic Patients. Cancer Practice, 6(5), 262-269. doi:10.1046/j.1523-5394.1998.00020.x

Juarez, G., Ferrell, B., &Borneman, T. (1999). Cultural Consideration for Education in Cancer Pain Management. *Journal of Cancer Education*, 14, 168-173. doi:10.1007/13187.1543-0154

Kosambi, D. D. (1967, February 1). Living Prehistory in India. *Scientific American,216*(2), 104-112. doi:10.1038/scientificamerican0267-104

Levine, J. D., Gordon, N. C., & Fields, H. L.. (1978). The Mechanism Of Placebo Analgesia.The Lancet, 312(8091), 654-57. Web. Retrieved June 28, 2016.

Lisb, T. L., Lund, I., Dahlin, L., Borg, E., Gustafsson, C., Sandin, L., . . . Eriksson, S. V. (2001). Reliability And Responsiveness Of Three Different Pain Assessments. *Journal of Rehabilitation Medicine,33*(6), 279-283. doi:10.1080/165019701753236473

Martinez, M. (2012, October 16). Endorphins. Retrieved June 28, 2016, from http://utpa-mmartinez23.blogspot.com/2012/10/endorphins.html

Melzack, R., & Wall, P. D. (1983). *The Challenge of Pain*. New York: Basic Books.

Morris, D. B. (1991). *The Culture of Pain*. Berkeley, CA: University of California Press.

Pascale, J. (n.d.). Alternative methods still important to Native healers. Retrieved June 17, 2016, from http://cojmc.unl.edu/nativedaughters/healers/alternative-methods-still-important-to-native-healers

"Placebo Effect." (2015, April 10). *American Cancer Society*. Web. Retrieved 29 June 2016.

Plotch, A. (2016). "The Placebo Effect: Is It Really Mind Over Matter?" BCA Chemistry. Retrieved July 18, 2016.

Riley, J. L., Wade, J. B., Myers, C. D., Sheffield, D., Papas, R. K., Price, D. D. (2002). Racial/ethnic Differences in the Experience of Chronic Pain. Pain, 100(3), 291-98. International Association for the Study of Pain. Web. Retrieved June 24, 2016.

Rotchford, J. K. (n.d.). Incorporating Medical Acupuncture into a Standard Medical Practice. Retrieved June 20, 2016, from http://www.medicalacupuncture.org/For-Patients/Articles-By-Physicians-About-Acupuncture/Incorporating-Medical-Acupuncture

Sumitra. (2013, January 28). Getting Stung by Bullet Ants, a Painful Initiation Ritual | Oddity Central - Collecting Oddities. Retrieved June 25, 2016, from http://www.odditycentral.com/pics/the-pain-of-growing-up-being-stung-by-hundreds-of-bullet-ants-in-the-amazon-rain-forest.html

"Synapse." (n.d.) *Medical Terms*. Web. Retrieved June 28, 2016.

"The Gate Control Theory of Chronic Pain." (2014, August 6). *Genesis Medical Clinic*. Web. Retrieved June 28, 2016.

"The Mechanism of Placebo Analgesia." (n.d.) *National Center for Biotechnology Information*. U.S. National Library of Medicine, Web. Retrieved June 29, 2016.

Thomas, V. J., and F. D. Rose. (1991) "Ethnic Differences in the Experience of Pain." *Social Science & Medicine* 32.9: 1063-066. Web. Retrieved June 25, 2016.

Thoms, Sue. (2015, April 19). "Grand Rapids Red Project: Saving Lives, One Overdose at a Time." *Michigan Live*. Web. Retrieved June 28, 2016.

"Understanding Naloxone." (n.d.). Harm Reduction Coalition. Web. Retrieved July 18, 2016,

What are opioids? (2014, November). Retrieved June 19, 2016, from https://www.drugabuse.gov/publications/research-reports/prescription-drugs/opioids/what-are-opioids

Wilner, A. N. (2008, October 14). Pain Management Across Cultures. Retrieved June 18, 2016, from http://www.medscape.org/viewarticle/581930

Woodruff, D. (2013, February 25). Adding Tempo Training to Run (and Race) Faster. Retrieved June 27, 2016, from http://www.blogher.com/adding-tempo-training-run-and-race-faster.

Locating Obnoxious Facilities: Minimizing Risk and Cost of Disposing and Transporting Hazardous Waste

Rahil Bathwal

Author background: Rahil Bathwal grew up in India and currently attends Jamnabai Narsee International School, located in Mumbai, India. His Pioneer seminar topic was in the field of mathematics and titled "Network Optimization: Facility Location Problems."

1. Introduction

While locating facilities, the primary goal is to ensure easy access for the entire population in the region. For both service facilities and emergency facilities, minimizing the distance between the facility and the people requiring the facility is usually the objective. However, in the case of obnoxious facilities, this is not valid. Unlike service or emergency facilities, obnoxious facilities constitute a class of facilities which are undesirable when located near people. It includes facilities like landfills and nuclear reactors which need to be located as far away as possible from settlements. Hence, the objective while locating obnoxious facilities is to maximize the sum of the vertex-weighted distances between demand nodes and facilities; therefore, this is a *maxisum* problem.

There are several other considerations when locating obnoxious facilities. Some of the key variables considered are the distance between the demand nodes and the facility, the cost of setting up new facilities, the transportation costs if they are applicable, the risk imposed by the facility, political and geographic factors and equity concerns i.e. sharing the burden of waste disposal amongst all waste producers. The first section of this paper deals with the objective of minimizing the distance, while the second section of this paper discusses an existing paper on this topic which deals with minimizing both risk to people and transportation costs. The third section of this paper develops a model to locate obnoxious facilities in a region of the state of Maharashtra in India.

1.1. Anti-medians and anti-centers

While choosing possible locations for obnoxious facilities we can consider measures of the anti-middle vertex such as anti-medians and anti-centers. In a network T, an *anti-center* is defined as a vertex with the maximum eccentricity. Therefore, the anti-center is the set of end vertices on the longest path(s) of the network T. In a network T, an *anti-median* is defined as a vertex with maximum status. An anti-median is a better location to position an obnoxious facility because it considers all the demand nodes and takes the average distance between them and a facility, while an anti-center only considers the longest path in the network. Therefore, the anti-median vertices are usually considered the optimal sites for obnoxious facilities. This section of the paper will use ideas and examples from *Centrality and Anticentrality in Trees*, a part of the *DIMACS Educational Module Series*.[1]

[1] Abueida, Atif, Michael Ackerman, and Sul-Young Choi. "Module 08-3: Centrality and Anticentrality in Trees." DIMACS Educational Module Series. Centre for Discrete Mathematics Theoretical Computer Science. DIMACS – Rutgers University. Centre for Discrete Mathematics Theoretical Computer Science, May, 2008. Web. 23 June 2016.

1.2. Anti-plurality vertices

Another possible measure for the anti-middle vertex of a tree is an anti-plurality vertex. In a tree T, V_{xy} represents the set of vertices closer to x than to y and $|V_{xy}|$ represents the number of such vertices. A vertex x is defined as an *anti-plurality vertex* only if $|V_{xy}| \leq |V_{yx}|$ for any given vertex y in the tree. This means that for any anti-plurality vertex x and for any given vertex y in the tree, there are never more vertices closer to vertex x than to vertex y. This ensures that the obnoxious facility is located at such a vertex x such that the minimum number of demand nodes are closer to this site in comparison to other potential sites for the facility.

***Proposition* 1.1.** An anti-plurality vertex is always an end vertex in a given tree T.
Proof. Let there be a vertex y that is an anti-plurality vertex, but not an end vertex of tree T. Let it be adjacent to vertices x and z. Consider $|V_{zy}|$ and $|V_{yx}|$. Vertex y is a part of V_{yx} but not V_{zy}. Now, let there be a vertex u that is closer to vertex z than to vertex y. Therefore, the unique simple path from vertex u to vertex y is the unique simple path from vertex u to vertex z with the additional edge zy. Furthermore, since T is a tree, the unique simple path from vertex u to vertex x must be the path from vertex u to vertex y with the additional edge xy. This implies that any vertex closer to vertex z than to vertex y is also closer to vertex y than to vertex x and therefore, $V_{zy} \subseteq V_{yx}$. Similarly, consider $|V_{xy}|$ and $|V_{yz}|$. Vertex y is a part of the V_{yz} but not V_{xy}. Now, let there be a vertex v that is closer to vertex x than to vertex y. This means that the unique simple path from vertex v to vertex y is the unique simple path from vertex u to vertex x with the additional edge xy. Furthermore, since T is a tree, the unique simple path from vertex v to vertex z must be the path from vertex v to vertex y with the additional edge zy. This implies that any vertex closer to vertex x than to vertex y is also closer to vertex y than to z and therefore, $V_{xy} \subseteq V_{yz}$. Furthermore, since vertex y is an anti-plurality vertex, the inequality $|V_{yx}| \leq |V_{xy}|$ must hold true. Therefore, the relationship among all four sets can be summarized by the inequality $|V_{zy}| \leq |V_{yx}| \leq |V_{xy}| \leq |V_{yz}|$. Unless all four sets are equal, this inequality violates the condition for y to be an anti-plurality vertex since for that to be true $|V_{yz}| \leq |V_{zy}|$ must also be true. However, the four sets cannot be equal since every vertex in V_{xy} is also in V_{yz} and V_{yz} contains the additional vertex y; similarly, every vertex in V_{zy} is also in V_{yx} and V_{yx} contains the additional vertex y. Therefore, the inequality is violated, proving that an anti-plurality vertex must also be an end vertex.

***Proposition* 1.2.** An anti-plurality vertex is always an end vertex on the longest path(s) in a given tree T.
Proof. The proof of this statement is similar to the proof of *Proposition 1.1* and can be found on page 15 of Module 08-3: *Centrality and Anticentrality in Trees*.[2]

<http://dimacs.rutgers.edu/Publications/Modules/Module08-3/dimacs08-3.pdf>
[2] Abueida, Atif, Michael Ackerman, and Sul-Young Choi. "Module 08-3: Centrality and Anticentrality in Trees." DIMACS Educational Module Series. Centre for Discrete Mathematics Theoretical Computer Science. DIMACS – Rutgers University. Centre for Discrete Mathematics Theoretical Computer Science, May, 2008. Web. 23 June 2016.
<http://dimacs.rutgers.edu/Publications/Modules/Module08-3/dimacs08-3.pdf>

1.3. Examples

***Example* 1.1.**

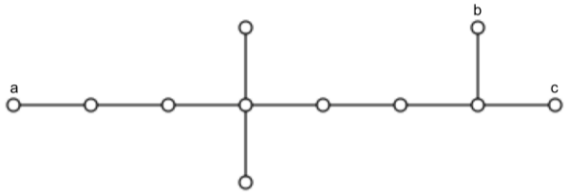


Figure 1

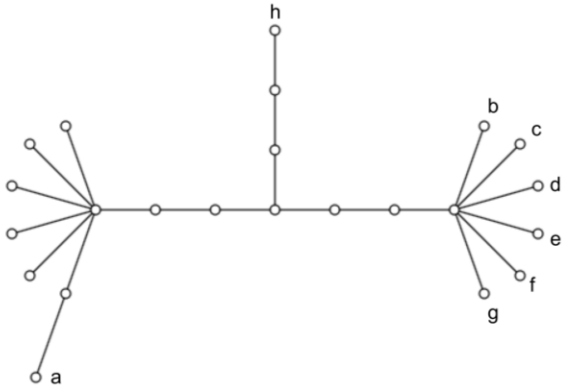


Figure 2

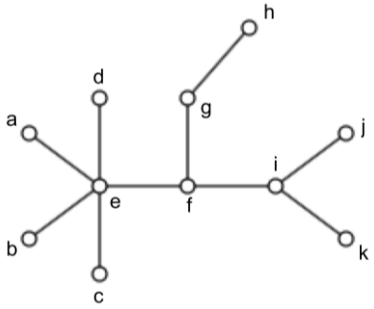


Figure 3

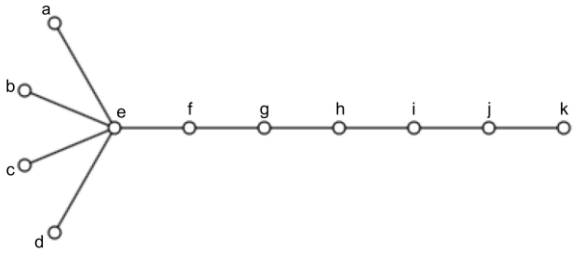


Figure 4

Question: Find the anti-plurality vertices, anti-medians and anti-centers of the graphs shown above.

Answer: In Fig. 1, the longest paths are $a - \cdots - b$ and $a - \cdots - c$. Thus, the anti-plurality vertices and anti-centers are a, b and c. However, the anti-median vertex is a since it has the maximum status. In Fig. 2, the longest paths are $a - \cdots - b, a - \cdots - c, a - \cdots - d, a - \cdots - e, a - \cdots - f$ and $a - \cdots - g$. Thus, the anti-plurality vertices and anti-centers are a, b, c, d, e, f and g. However, the anti-median vertex is h since it has the maximum status.

In Fig. 3, the longest paths are $a - \cdots - h, a - \cdots - j, a - \cdots - k, b - \cdots - h, b - \cdots - j, b - \cdots - k, c - \cdots - h, c - \cdots - j, c - \cdots - k, d - \cdots - h, d - \cdots - j, d - \cdots - k, h - \cdots - j$ and $h - \cdots - k$. Thus, the anti-plurality vertices and anti-centers are a, b, c, d, h, j and k. However, the anti-median vertex is h since it has the maximum status. In Fig. 4, the longest paths are $a - \cdots - k, b - \cdots - k, c - \cdots - k$ and $d - \cdots - k$. Thus, the anti-plurality vertices and anti-centers are a, b, c, d and k. However, the anti-median vertex is k since it has the maximum status.

***Example* 1.2.**

Question: Find a tree for which the anti-median, anti-plurality vertices and anti-centers are the same vertices.

Answer: In the tree shown below, the anti-plurality vertices, anti-medians and anti-centers are the four end vertices.

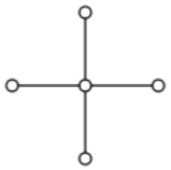


Figure 5

***Example* 1.3.**

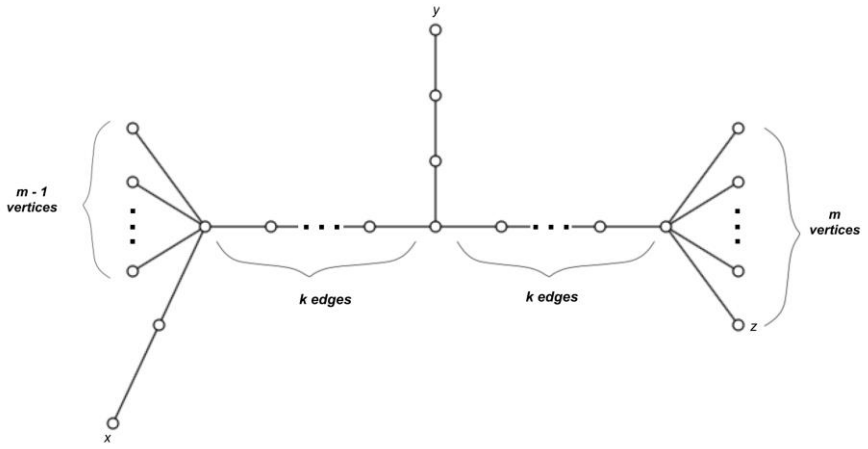


Figure 6

Question: For the given tree, find the value of k and m for which the distance between any anti-median and any anti-plurality vertex (or any anti-center) is at least 9.
Answer: The longest path in the tree is $x - \cdots - y$ or $x - \cdots - z$, depending on the chosen value of k. The $x - \cdots - y$ path is of length $k + 5$ and the $x - \cdots - z$ path is of length $2k + 3$. The minimum value of k for which any of these paths is of length 9, is $k = 3$ because when $k = 3, d(x, y) = 8$ and $d(x, z) = 9$. Therefore, the anti-plurality vertices will be x and z as well as all the $m - 1$ vertices above z. For the distance between the anti-median and anti-plurality vertex to be 9, vertex y must be the anti-median vertex. Therefore, $d(x, y)$ and $d(y, z)$ must both be at least 9, which means $k + 5$ and $k + 4$ must both be at least 9, therefore, k must be at least 5. If the graph is drawn with $k = 5$, then a suitable value of m needs to be chosen such that the status of vertex y is the maximum. For $k = 5$, the status of vertex x is $102 + 16m$, the status of vertex y is $76 + 18m$ and the status of vertex z (and all the similar $m - 1$ vertices above z) is $101 + 14m$. Therefore, for vertex y to have the maximum status, the inequality $76 + 18m \geq 102 + 16m$ must hold true. When this inequality is solved, the solution obtained is $m > 13$. Therefore, $m \geq 14$ and the first solution to the question is $k = 5$ and $m = 14$.

1.4. More realistic models

The methods discussed above focus on choosing the optimal site for locating obnoxious facilities using the properties of trees. However, in a real-world scenario, the underlying network of settlements may not be a tree and will have weighted edges (due to the different lengths of roads) and weighted vertices (due to the populations of settlements). Several sites may be almost equally optimal, and there are other important factors to consider. If the total distance between demand nodes and facilities is maximized, the transportation costs can increase, and this might make a certain site less optimal. Hence, when deciding the location for an obnoxious facility, minimizing both the risk and transportation costs becomes an important objective. Models used in existing literature focus on either one of these objectives or combine them into a single formulation.

2. Literature Review

This section will discuss a multi-objective model developed by Siber Alumur and Bahar Y. Bara in their paper titled *A New Model for the Hazardous Waste Location-Routing Problem*.[3] The paper develops a model based on the predetermined candidate sites for waste treatment and waste disposal facilities and then uses a multi-objective function to identify the optimal candidate sites as well as the number of facilities to be built. It addresses the problem of locating both waste treatment and disposal centers and applies the model to the Central Anatolian region of Turkey.

2.1. Review of a current model

The two primary objectives of the model developed by Alumur and Bara are to minimize the risk during transportation and the costs of transportation. The model incorporates waste being transported from a generation node to a treatment facility and the residual waste from the facility then being transported to a disposal facility. It also allows for the possibility that a generation node, treatment facility and disposal facility are situated at the same site. Furthermore, it also incorporates different waste types and different treatment types into the model.

[3] Alumur, Sibel, and Bahar Y. Kara. "A New Model for the Hazardous Waste Location-Routing Problem." Computers Operations Research 34.5 (2007): 1406-423.

2.2. Multi-objective mixed integer programming formulation

Alumur and Bara's model defines $N = (V, A)$ as the transportation network and uses $G = \{1, \ldots, g\}$ as the set of all generation nodes, $T = \{1, \ldots, t\}$ as the set of possible sites for treatment nodes, $D = \{1, \ldots, d\}$ as the set of possible sites for disposal nodes, $Tr = \{1, \ldots, tr\}$ as the set of possible sites for transshipment nodes, $W = \{1, \ldots, w\}$ as the set of possible waste types and $Q = \{1, \ldots, q\}$ as the set of possible treatment technologies such as incineration and chemical treatment. The model defines various parameters including $c_{i,j}$ and $cz_{i,j}$ as the costs of transporting hazardous waste and waste residue on link $(i,j) \epsilon A$ respectively; the cost of opening a treatment facility and disposal facility at node i as $fc_{q,i}$ and fd_i respectively; the amount of waste generated by node i as $g_{w,i}$ and percent mass reduction of this waste after treatment as $r_{w,q}$; capacity of treatment and disposal facilities as $t_{q,i}$ and dc_i respectively; and POP_{wij} as the number of people affected by the transportation of hazardous waste along link (i,j).

The model is a mixed integer program with both real and binary decision variables. The real decision variables incorporated in the model are the amount of waste of type w that is shipped from a generation node to a treatment facility through link (i,j) as $x_{w,i,j}$; the amount of residual waste shipped from a treatment facility to a disposal facility through link (i,j) as $z_{i,j}$; the amount of waste of type w treated at a given treatment facility as $y_{w,q,i}$, and the amount of waste disposed of at a disposal facility as d_i. As stated by Alumur and Bara, the number of real decision variables depends on the number of waste types (w), links (m), treatment nodes (t), disposal nodes (d) and treatment technologies (q) and is equal to $(wm + m + qt + d)$. The binary decision variables included represent whether or not a treatment or disposal facility is built at a candidate site as $f_{q,i}$ and dz_i. The decision variable is 1 if the site is part of the optimal solution and 0 if not. As stated by Alumur and Bara, the number of binary decision variables depends on the number of treatment technologies (q), treatment nodes (t) and disposal nodes (d) and is equal to $(qt + d)$.

The formulation used by Alumur and Bara in their paper minimizes the function

$$\sum_{(i,j)\epsilon A} \sum_{w} c_{i,j} x_{w,i,j} + \sum_{(i,j)\epsilon A} cz_{i,j} z_{i,j} + \sum_{i} \sum_{q} fc_{q,i} f_{q,i}$$

to minimize cost and simultaneously minimizes the function

$$\sum_{(i,j)\epsilon A} \sum_{w} POP_{w,i,j} x_{w,i,j}$$

to minimize risk. The first function represents the transportation and initial setup costs by considering the amount of waste transported through a given link multiplied by the cost of transporting it through that link for all waste types and all such links $(i,j) \epsilon A$, the amount of residual waste transported through a given link multiplied by the cost of transporting it through all such links links $(i,j) \epsilon A$, and the cost of opening a treatment facility at node i multiplied by whether or not that site is chosen in the optimal solution; this is done for all treatment technologies and candidate nodes. The second function assesses the risk posed by the facility by considering the number of people affected by a certain waste type carried through the link (i,j) and multiplying it by the amount of waste transported through that link. This is done for all waste types and all links (i,j).

The feasible solutions are defined by various constraints. The two important constraints used are mass and flow balance. The first of these constraints is the flow balance constraint, which is defined by the equation

$$(1 - \alpha_{w,i}) g_{w,i} = \sum_{j:(i,j)\epsilon A} x_{w,i,j} - \sum_{i:(i,j)\epsilon A} x_{w,i,j} + \sum_{q} y_{w,q,i} \ , w \epsilon W, i \epsilon V.$$

This constraint ensures that the total amount of waste generated at a given node that is not recycled is equal to the net flow of waste from that node and the amount of waste treated at that node. The second constraint used is the mass and flow balance constraint, which is defined by the equation

$$\sum_{q}\sum_{w} y_{w,q,i}(1 - r_{w,q})(1 - \beta_{w,q}) - d_i = \sum_{j:(i,j)\epsilon A} z_{i,j} - \sum_{j:(i,j)\epsilon A} z_{j,i} \ , i \epsilon V.$$

This constraint ensures that the total amount of residual waste that cannot be treated is equal to the net flow of disposable residual waste to a node with a disposal facility. The other important constraints are the capacity constraints for the located facilities, which can be represented by the relations

$$\sum_{w} y_{w,q,i} \leq t_{q,i} f_{q,i} \quad and \quad d_i \leq dc_i dz_i.$$

These constraints ensure that that the amount of waste of a given type shipped to a treatment facility or a disposal facility is less than the amount of waste that can be processed at that facility multiplied by whether or not the facility is built on that site in the optimal solution. The two other important constraints in this model ensure that waste is not treated or disposed at an inappropriate facility, i.e. a site that is not a candidate site for the problem. These constraints are

$$\sum_{q}\sum_{w} y_{w,q,i} = 0, i \epsilon (V - T), q \epsilon Q, i \epsilon T \quad and \quad d_i = 0, i \epsilon (V - D), i \epsilon D.$$

These constraints ensure that once we subtract the set containing candidate sites from the set containing all nodes, the remaining nodes will have no waste treated or disposed of at them.

As stated by Alumur and Bara, the number of constraints depends on the number of waste types (w), nodes (n), treatment nodes (t), disposal nodes (d) and treatment technologies (q). Thus, the number of constraints is $(wn + n + 2qt + d + wqt + (n - t) + (n - d))$. Therefore, when this model is used to solve a facility location problem, where the number of candidate sites increases and the waste produced has a wide range, the model has to consider a much larger number of variables and constraints as compared to the application in Turkey. Hence, the number of feasible solutions increases exponentially as the size of the problem grows.

Since this model is multi-objective and tries to minimize two functions, Alumur and Bara have used a weighted approach to solving facility location problems. Using λ to represent the weighted importance given to minimizing cost, they have developed a convex combination of risk and cost, which is of the form

$$\lambda \times \left(\frac{cost\ of\ link}{maximum\ cost\ of\ link}\right) + (1 - \lambda) \times \left(\frac{risk\ of\ link}{maximum\ risk\ of\ link}\right) \ with\ \lambda \epsilon [0, 1].$$

The cost and risk have been divided by the maximum cost and risk respectively to scalarize the function. Since the cost and risk are measured differently, they cannot be directly combined into a single multi-objective function; therefore, this method of scalarization has been chosen. However, this method is limited in the instances where the data for a given problem has a wide range, in which case the maximum cost or risk could skew the data to one side; since the given application ignores this possibility and uses a small range of data, the model cannot be applied to larger scenarios. Using this weighted multi-objective approach for the application in Turkey, Alumur and Bara suggested that the best values for λ are between 0.2 and 0.7. However, their data and trade-off curve show that the best compromise is when $\lambda = 0.6$ or 0.7. Choosing a higher or lower value for λ gives too

much weight to risk or cost.

The model developed by Alumur and Bara was implemented in Turkey for two scenarios: one with 15 candidate sites and one with 20 candidate sites. While the optimal solutions obtained from both scenarios were fairly similar, the computation time required by CPLEX to solve the problem using this model increased when repeated for 20 candidate sites. While the time required for solving the problem with 15 candidate sites ranged between 0.33 and 2.26 hours depending on the case being tested, the problem with 20 sites required times ranging between 0.71 and 16.33 hours. This exponential increase in time when the problem gets larger is a major limitation for this model. Hence, for larger problems, heuristic methods may need to be employed.

2.3. Efficiency and accuracy of this model

The model discussed above assumes complete knowledge about the amount of waste produced at a given generation node. However, when the model was implemented in Turkey, there was insufficient data about how much hazardous waste was produced by each district of Turkey, and estimates had to be used which were based on evaluating the population of a given district and the level of industrial activity conducted in the region. Furthermore, this model also assumes complete knowledge of the costs involved in opening a new treatment or disposal facility. However, as discussed by Alumur and Bara in their paper, there are several factors like building material prices, taxes and land prices which affect the costs of building a new facility, and this data was not available to them. Thus, estimates had to be used for the model. Another practical problem with this model is determining the candidate sites which the model will test. Alumur and Bara used various methods, possibly including the methods discussed in Section 1 of this paper about the anti-middle of a graph, to determine the potential sites. However, if such a model were used in the real world to solve a facility location problem, the candidate sites would be chosen by authorities such as the government and this would bring in several non-mathematical considerations which this model cannot accommodate.

Another limitation of the model is that if there are more than 20 candidate sites, the time required to find the optimal solutions increases exponentially, even when powerful software is used for calculations, which signals the need for the development of heuristic methods. The biggest limitation of the model is its ability to evaluate all the risks associated with obnoxious facilities; it only considers the risks of transporting waste through a given link and does not take into account the risk of situating a facility at a given location. It also fails to consider equity, i.e. people bearing the burden of the waste equally.

However, the model is an effective formulation because it considers variables not incorporated in models formulated in other existing literature. Moreover, it provides routing strategies which supply more comprehensive data for implementing the solutions obtained from this model. Also, this model has been applied to a problem of a much larger scale compared to other papers on the same topic.

3. Application in India

Waste management is an area of concern in India since it has a large population and hence the amount of waste generated is very high. It is necessary to find appropriate means of disposing of this waste to prevent any negative repercussions on the health of the Indian population. This makes it very important to build landfills at appropriate locations and requires the evaluation of several variables and objectives. This section will focus on locating landfills in the state of Maharashtra, which produces very large amounts of waste due to metropolitan cities like Mumbai.

3.1. Defining the problem scope

Since the state of Maharashtra is very large, covering an area of 307, 713 km^2, it is not feasible to consider the entire state in a single model while locating landfills.[4] Hence, to solve a more reasonable problem, this paper considers a certain set of districts in the southwest part of the state. The map in Fig. 7 and Table 1 shown below highlight the districts considered in the state.[5] The data for the districts on the map was obtained from the Government of Maharashtra's website.[6]

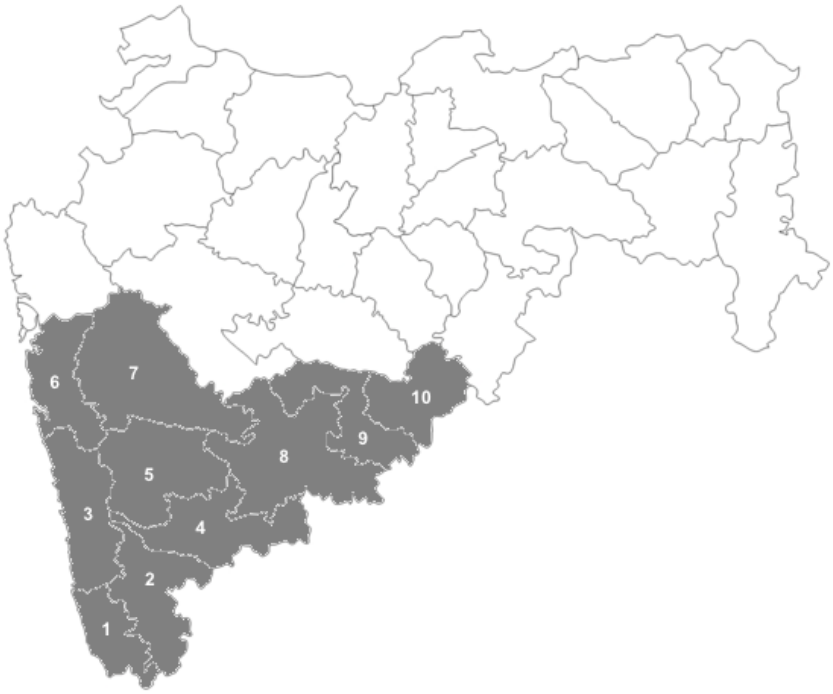


Figure 7: *Chosen Districts in Maharashtra*

[4] Mission Ministry of Health Family Welfare. "State Wise Information: Maharashtra." National Health Mission, Ministry of Health Family Welfare, Government of India. Government of India Ministry of Health Family Welfare, n.d. Web. 20 June 2016. <http://nrhm.gov.in/nrhm-in-state/state-wise-information/maharashtra.html>

[5] D-maps.com. "Maharashtra Outline, Districts (white)." D-maps.com. D-maps.com, n.d. Web. 19 June 2016. <http://www.d-maps.com/carte.php?num_car=31837&lang=en>

[6] Government of Maharashtra. "Districts." Official Website of Government of Maharashtra. Government of Maharashtra, n.d. Web. 19 June 2016. <https://www.maharashtra.gov.in/>

Number	District
1	Sindhudurg
2	Kolhapur
3	Ratnagiri
4	Sangli
5	Satara
6	Raigad
7	Pune
8	Solapur
9	Osmanabad
10	Latur

Table 1: *Chosen Districts in Maharashtra*

3.2. Problem formulation

To identify the optimal location of the landfills, a two-step approach was chosen. One primary issue of concern while locating landfills is equity. No community wants to bear the entire burden of everybody's waste. Hence, to address this concern, one possible candidate site was identified in each district. This ensures that each district has a possible site for the landfill.

Another important objective while locating obnoxious facilities like landfills is to maximize the distance between the demand nodes and the closest facility. If a facility was built in a particular district, it would serve all the settlements in that district. Hence, the candidate site for each district can be chosen based on two objectives. The two options are to choose either the median or the anti-median of the district with both options satisfying different objectives. Choosing the median reduces the average distance between a demand node and the facility, which lowers the transportation cost of shipping the waste. On the other hand, choosing the anti-median increases the average distance between a demand node and the facility. This is considered desirable because people want landfills to be as far away as possible from their places of residence. Moreover, this also reduces the risk of the waste having a negative effect on the people's health. Since the areas of the districts in question are not very large, transportation costs would not differ too much within a single district; however, they would become a greater concern if a facility were to serve a larger area. Hence, to choose the candidate site for each district, the anti-median was found. This was done because minimizing risk seemed more important than minimizing transportation costs due to the small size of each district.

As defined in Section 1.1, the anti-median of a network is the node with the largest status. Since each district's population is spread across the district, and it is not possible to create clear demarcations, the five most populous towns of each district were identified. The status of each was computed, and the site with the largest status was identified as the district's anti-median. This was done for each district to identify ten candidate sites. The problem of finding the anti-median of a network can be formulated as a binary integer program. The following model is adapted from the classic *p*-median problem formulation, as

shown in *Network and Discrete Location* by Mark Daskin.[7] All the population data was obtained from the Government of India's 2011 census publications.[8] All the distance data was obtained from Google Maps.[9]

The set of five most populous sites of a district form the transportation network $N = (V, E)$ and the status of each site is given by $s(v)$. The following notation is defined for this network.

Inputs

$p(u) =$ population of node $u \in V$;

$d(u, v) =$ distance between demand node $u \in V$ and a facility node at $v \in V$.

Decision variables

$X_v = 1$ if we do locate landfill and 0 if we do not at site $v \in V$.

Using the notation defined above, the following model is devised for this network.

$$\text{Maximize:} \qquad \sum_{v \in V} s(v) X_v = \sum_{v \in V} \left(\sum_{u \in V} p(u) d(u, v) \right) X_v$$

$$\text{Subject to:} \qquad \sum_{v \in V} X_v = 1$$

$$\text{with:} \qquad X_v \in \{0, 1\} \; for \; all \; v \in V$$

The constraint $\sum_{v \in V} X_v = 1$ ensures that a landfill is located at only one of the five possible sites since we only need one anti-median for each district. The solution obtained for each district was the set of anti-medians for that district. However, not all solutions are geographically viable since Maharashtra has mountainous regions and some of the solutions sites were hill stations. These sites were the anti-medians of their districts but they were not appropriate sites for landfills. Furthermore, some solution sites were military camps, which are not viable for landfills since they are not politically preferable. Hence, geographic and political factors were also incorporated by eliminating sites due to these non-mathematical reasons, and choosing the site with the second highest status as the anti-median for that district. The anti-medians for each district are shown in Table 2.

[7] Daskin, Mark S. Network and Discrete Location: Models, Algorithms, and Applications. 2nd ed. Hoboken: John Wiley Sons, 2013. Wiley Online Library. John Wiley Sons, Inc, 25 June 2013. Web. 23 June 2016. <http://onlinelibrary.wiley.com/book/10.1002/9781118537015>

[8] Office of the Registrar General Census Commissioner. "District Census Hand Book - Maharashtra." Census of India Website: Office of the Registrar General Census Commissioner, India. Government of India Ministry of Home Affairs, n.d. Web. 20 June 2016. <http://www.censusindia.gov.in/2011census/dchb/Maharashtra.html>

[9] Google. "Google Maps." Google Maps. Google, n.d. Web. 20 June 2016. <https://www.google.co.in/maps>

District	Anti-Median	Site Number
Sindhudurg	Kankavli	1
Kolhapur	Jaysingpur	2
Ratnagiri	Dapoli Camp	3
Sangli	Vita	4
Satara	Phaltan	5
Raigad	Khopoli	6
Pune	Pimpri Chinchwad	7
Solapur	Akluj	8
Osmanabad	Paranda	9
Latur	Ahmadpur	10

Table 2: *Anti-Medians for Each District*

Having identified 10 candidate sites for the landfill, the next step is to choose the optimal site for the landfill. Each candidate site is representative of its district. Hence, the demand of each site will be the demand of the entire district and not just the site itself because if a facility is built at that site, the entire district will ship its waste to that facility. The two primary objectives for locating landfills are minimizing transportation costs and maximizing distance between people and facilities. In the first step, the distance between facilities and people was maximized by choosing the anti-median of each district. However, in the second step, when there were ten districts under consideration, the transportation costs could have increased dramatically if the sites were chosen based on the objective of maximizing distance between settlements and facilities. Hence, a different approach has been used for the second step. Two objectives have been considered: minimizing transportation costs and minimizing risk from people living in close proximity to a facility. This is a multi-objective approach to the problem and hence requires a combination of the two objectives.

The first objective is to minimize transportation cost and the variables which affect this are the distance between the demand node and facilities, the amount of waste being transported and the cost of shipping waste between the two locations. The cost of shipping the waste is proportional to the distance between two nodes and the amount of waste being transported between the two nodes since the cost will be a fixed constant per kilometer per gram depending on fuel, vehicle and labor prices. Hence, by incorporating the distances between the demand nodes and facility nodes and the amount of waste transported, the model also considers the cost incurred. The model also accounts for the industrial activity in the demand node, which affects the overall waste produced per capita. According to the 1995 survey, the data shown below in Table 3 was obtained.[10]

[10] Asnani, P. U. “Solid Waste Management.” Ed. A. Rastogi. India Infrastructure Report 2006 Urban Infrastructure (2006): 160-89.

Population range (in millions)	Average per capita waste generation per day (in grams)
0.1 to 0.5	210
0.5 to 1.0	250
1.0 to 2.0	270
2.0 to 5.0	350
5.0 plus	500

Table 3: *Waste Generation per capita in Indian cities*

As mentioned previously, the demand of the entire district will be attributed to the anti-median of that district for the second step. To compute the demand of the entire district, the total amount of waste generated needs to be computed. This was done by using the data in the table above and the total population of all the settlements of the given district to find the total amount of waste generated in the district. For a city with population z, the total waste production will be $P(z) = w(z) \times z$, where $w(z)$ is the per capita waste production, as given in Table 3. Using this notation, if a district has y cities of populations $z_1, z_2, z_3, \dots, z_y$, then the total waste produced by that district is given by $P(z_1) + P(z_2) + P(z_3) + \cdots + P(z_y)$. Using this method, the total waste produced was computed for each district and is shown in Table 4.

District	Total Waste Generated per Day (ton)
Sindhudurg	178.43
Kolhapur	835.93
Ratnagiri	339.16
Sangli	612.76
Satara	630.79
Raigad	553.18
Pune	2521.26
Solapur	944.79
Osmanabad	348.09
Latur	515.38

Table 4: *Total Waste Production for Each District*

The objective of this second step is to minimize the transportation cost and risk due to proximity to a facility. The risk posed by the facility primarily affects the people living in the city chosen to build the landfill. Hence, the function to minimize the risk due to proximity to a landfill incorporates the population of the candidate sites of each district and not the entire district's population. For this model, the following notation is defined.
$N = (V, E)$ transportation network;

$WP = \{1, \dots, wp\}$ waste production nodes;
$L = \{1, \dots, l\}$ landfill candidate sites.
Inputs
$p(v) =$ population of node $v \epsilon L$;
$d(u, v) =$ distance between demand node $u \epsilon WP$ and a facility at node $v \epsilon L$;
$g_u =$ amount of waste generated at node u;
$h =$ number of landfills to be built.
Decision Variables
$t_v =$ amount of waste disposed at landfill at node v;
$X_v = 1$ if we do locate a landfill and 0 if we do not at site $v \epsilon L$;
$Y_{u,v} = 1$ if waste generation node u is served by a landfill at node v and 0 if it is not.
Using the above notation, the following mathematical model has been developed.

Minimize: $$\sum_{v \epsilon L} \left(\sum_{u \epsilon WP} g_u d(u, v) Y_{u,v} \right) X_v$$

AND

Minimize: $$\sum_{v \epsilon V} p(v) t_v X_v$$

Subject to:

$$\sum_{v \epsilon V} X_v = h \quad (1)$$

$$\sum_{u \epsilon WP} Y_{u,v} \leq 10 X_v \; for \; all \; v \epsilon L \quad (2)$$

$$\sum_{u \epsilon WP} g_u Y_{u,v} = t_v \; for \; all \; v \epsilon L \quad (3)$$

$$\sum_{v \epsilon L} Y_{u,v} = 1 \; for \; all \; u \epsilon WP \quad (4)$$

$$t_v \geq X_v \quad (5)$$

with:

$$X_v \epsilon \{0, 1\} \; for \; all \; v \epsilon V$$

$$Y_{u,v} \epsilon \{0, 1\} \; for \; all \; u \epsilon WP, v \epsilon L$$

$$t_v \geq 0$$

The first function minimizes the transportation cost by minimizing the distance the waste produced by any node $u \epsilon WP$ is shipped to be disposed of at the closest landfill at node $v \epsilon L$. The function takes into consideration whether the waste from node u is shipped to a landfill at node v and whether or not there is a landfill at node v. The second function minimizes the risk of the hazardous waste at a landfill at node v by minimizing the amount of waste processed at a landfill at node v multiplied by the population of node v. This function takes into consideration whether the node v is a solution site or not. The first constraint ensures that the number of landfills built is equal to the input for h. The second constraint ensures that a waste generation node u does not ship its waste to node v unless node v has a landfill. The third constraint ensures that the total amount of waste disposed of

at a landfill at site v is equal to the total amount of waste shipped from all the waste production nodes u that are served by the landfill at node v. The fourth constraint ensures that each waste generation node u is assigned to exactly one facility. Finally, the fifth constraint ensures that if a landfill is opened at a candidate site then at least some waste is disposed of there.

Since the model uses a multi-objective approach, the two functions need to be combined. Similar to the approach taken by Alumur and Bara, a convex combination has been chosen. The two variables, transportation cost and risk to residents of the district, have to be scalarized. In this paper, the variables have been divided by their minimums to scalarize them. The following solution technique is used.

$$\lambda \times \left(\frac{cost\ of\ link}{minimum\ cost\ of\ link}\right) + (1-\lambda) \times \left(\frac{risk\ of\ link}{minimum\ risk\ of\ link}\right)$$

Using the model and the solution approach suggested above, the formulation was applied in the chosen districts of Maharashtra. Excel Solver was used to find the optimal solutions. There were five cases evaluated with h = 1, 2, 3, 4 and 5 respectively. This was done for varying values of λ from 0 to 1 with increments of 0.25. The tables below summarize the results obtained.

Problem	**Solution sites**				
	h = 1	h = 2	h = 3	h = 4	h = 5
λ = 0.00	3	3, 1	1, 3, 9	1, 3, 8, 9	1, 3 8, 9, 10
λ = 0.25	3	3, 9	1, 3, 9	1, 3, 9, 10	1, 3, 4, 9, 10
λ = 0.50	9	3, 9	2, 3, 9	2, 3, 6, 9	2, 3, 6, 8, 10
λ = 0.75	9	6, 8	2, 6, 8	2, 6, 8, 10	2, 3, 6, 8, 10
λ = 1.00	7	7, 8	2, 7, 8	2, 7, 8, 10	2, 3, 7, 8, 10

Table 5: *Summary of Solution Sites for Each Case*

Problem	**Total Risk (ton people) / 10^8**				
	h = 1	h = 2	h = 3	h = 4	h = 5
λ = 0.00	1.18	1.18	1.18	1.23	1.32
λ = 0.25	1.18	1.25	1.26	1.39	1.59
λ = 0.50	1.40	1.28	1.72	3.49	4.03
λ = 0.75	1.40	4.05	4.19	4.21	4.03
λ = 1.00	129.23	60.60	60.74	60.76	54.95

Table 6: *Summary of Total Risk for Each Case*

Problem	Total Cost (ton people) / 10^6				
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
$\lambda = 0.00$	1.75	1.71	1.85	1.88	1.76
$\lambda = 0.25$	1.75	1.24	1.11	1.01	0.99
$\lambda = 0.50$	1.43	1.22	0.96	0.50	0.31
$\lambda = 0.75$	1.43	0.68	0.50	0.36	0.31
$\lambda = 1.00$	1.06	0.56	0.38	0.25	0.18

Table 7: *Summary of Total Cost for Each Case*

As can be seen in the tables above, the case for $\lambda = 1.00$ gives an extreme value of risk because it locates a facility at Pimpri Chinchwad, which has a very large population compared to the other potential sites. Hence, although the cost is minimized, the risk is increased dramatically and this makes this scenario undesirable. The other four values for λ are all possible approaches with acceptable solutions. However, to find the best possible value of λ, a trade-off curve was plotted between cost and risk for the case $h = 3$ by varying the value of λ with increments of 0.1. This approach is similar to the approach taken by Alumur and Bara in their paper. Table 8 shown below shows the values of cost and risk for the different values of λ.

λ	Total Cost (ton km) / 10^6	Total Risk (ton people) / 10^8
0	1.85	1.18
0.1	1.15	1.24
0.2	1.11	1.26
0.3	1.11	1.26
0.4	1.11	1.26
0.5	0.96	1.72
0.6	0.95	1.78
0.7	0.56	3.68
0.8	0.50	4.19
0.9	0.50	4.19
1.0	0.38	60.7

Table 8: *Cost and Risk for $h = 3$*

As can be seen in Table 8, certain values of λ give the same solution and value for risk and cost. Therefore, some data points on the following trade-off curve correspond to multiple values of λ. The trade-off curve does not include Table 8's last data point since the value for risk is too extreme and skews the rest of the data.

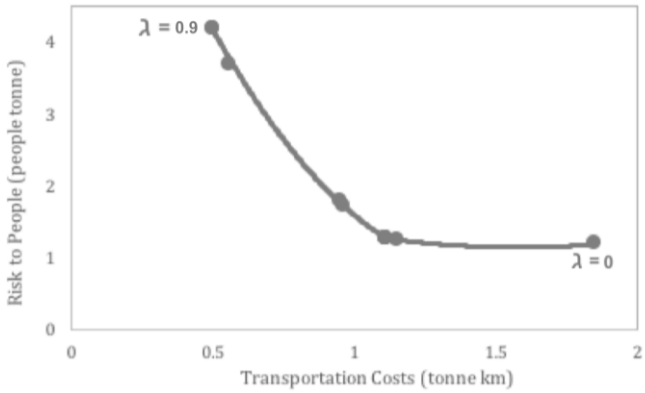


Figure 8: *Risk and Cost Trade-Off Curve for h = 3*

As can be seen from the trade-off curve and the data in Table 8, the best values for λ seem to be between 0.1 to 0.6. This range shows that giving greater importance to minimizing risk results in an overall better compromise between risk and cost. This shows that for the state of Maharashtra and the 10 districts considered in this model, risk has a greater effect when deciding the best location for landfills. Also, as the data shows, when the value of λ is at its extreme, i.e. 0 or 1, the solutions are not feasible. When $\lambda = 0$, the transportation cost increases considerably and when $\lambda = 1$ the risk increases considerably. However, the increase is more pronounced for $\lambda = 1$.

3.3. Conclusion

The model developed for the application in India considers multiple variables and provides solutions for five cases - 1, 2, 3, 4 and 5 landfills with different weights given to two objectives. However, the number of landfills to be built is a decision that depends on the space available for a landfill, the cost of building a landfill and other factors. This model does not consider these variables and the ultimate decision is left to the deciding authorities depending on the specific requirements for the solution. Another important aspect of this model is that it assumes that there is only one kind of waste and that the waste is not treated or recycled before being disposed of. Depending on the kind of waste produced and the required treatments, this model will have to be modified to incorporate different waste types as well as treatment and recycling facilities as an intermediate step before disposal.

The formulation used in this paper is a two-step process. The first step is to find the set of potential locations and then the second step is to choose the optimal landfill sites. The advantage of this method is that it does not make the selection of candidate sites an arbitrary choice. In fact, it provides a mathematical model reliant on anti-medians to compute the best set of candidate sites for the landfills. Moreover, solution sites which are geographically or politically inviable were eliminated at this stage. However, there are various other political factors which can be incorporated when such a model is implemented by the authorities. Further research could include more towns from each district instead of including just the five most populous ones considered in this paper. This is an advantage over a method reliant on arbitrary selection of candidate sites because those sites may not be the best-suited sites for landfills in their respective districts. Another notable aspect is that the data used from the 1995 research linking population levels to the per capita waste production is 21 years old at the time of writing this paper. It is possible that, on a relative scale, the population and waste production rate in each town would have grown at roughly an equal rate and hence this

would not affect the solutions provided in this paper. However, it would affect the value of the objective function and the values of cost and risk would change. Similarly, the population data used is from the 2011 Census in India (the next being in 2021) and hence, the population numbers are five years old. Therefore, the old data does not affect the solutions, but it does affect the cost and risk which would be important for governments and authorities when making final decisions. Nonetheless, if the most recent data is implemented into the model, the formulation yields accurate results.

This model is flexible since it allows for the possibility of accommodating existing landfills by setting the value of X_v to 1 for the concerned candidate site. Hence, this model can be applied to regions with existing landfills, which are not looking for a new system but only hoping to build more landfills due to increasing demand. The multi-objective approach used in this model considers the transportation cost as well as the risk to the population a facility poses when located at a given site. The multi-objective solution technique can be adapted as per the requirements of the region. In the chosen application in Maharashtra, the risk to people was very high when the transportation costs were minimized. Hence, choosing maximum cost and risk link instead of minimum cost and risk link would skew the data in the scalarization process. However, for other regions, other techniques might prove to be more reasonable and this model allows for that flexibility. Furthermore, this model accounts for the risk of locating a facility at a given site, a variable not included in the formulation developed by Alumur and Bara. This variable is very important in the Indian context due to the large population density of certain cities. Hence, although this model lacks certain variables incorporated in Alumur and Bara's model, it includes a very important variable i.e. risk of locating a facility at a given location and hence is adapted for the application in India. Further use of this model could increase the size of the problem, perhaps due to a greater number of districts or candidate sites in each district. In such a scenario, Excel Solver would not be an appropriate software and implementing the model would require higher grade optimization software and perhaps heuristics since calculation times will increase exponentially if more variables are included.

References

Abueida, Atif, Michael Ackerman, and Sul-Young Choi. "Module 08-3: Centrality and Anticentrality in Trees." DIMACS Educational Module Series. Centre for Discrete Mathematics Theoretical Computer Science. DIMACS – Rutgers University. Centre for Discrete Mathematics Theoretical Computer Science, May, 2008. Web. 23 June 2016. <http://dimacs.rutgers.edu/Publications/Modules/Module08-3/dimacs08-3.pdf>

Alumur, Sibel, and Bahar Y. Kara. "A New Model for the Hazardous Waste Location-Routing Problem." Computers Operations Research 34.5 (2007): 1406-423.

Asnani, P. U. "Solid Waste Management." Ed. A. Rastogi. India Infrastructure Report 2006 Urban Infrastructure (2006): 160-89.

Daskin, Mark S. Network and Discrete Location: Models, Algorithms, and Applications. 2nd ed. Hoboken: John Wiley Sons, 2013. Wiley Online Library. John Wiley Sons, Inc, 25 June 2013. Web. 23 June 2016. <http://onlinelibrary.wiley.com/book/10.1002/9781118537015>

D-maps.com. "Maharashtra Outline, Districts (white)." D-maps.com. D-maps.com, n.d. Web. 19 June 2016. <http://www.d-maps.com/carte.php?num_car=31837&lang=en>

Google. "Google Maps." Google Maps. Google, n.d. Web. 20 June 2016. <https://www.google.co.in/maps>

Government of Maharashtra. "Districts." Official Website of Government of Maharashtra. Government of Maharashtra, n.d. Web. 19 June 2016. <https://www.maharashtra.gov.in/>

Mission Ministry of Health Family Welfare. "State Wise Information: Maharashtra." National Health Mission, Ministry of Health Family Welfare, Government of India. Government of India Ministry of Health Family Welfare, n.d. Web. 20 June 2016. <http://nrhm.gov.in/nrhm-in-state/state-wise-information/maharashtra.html>

Office of the Registrar General Census Commissioner. "District Census Hand Book - Maharashtra." Census of India Website: Office of the Registrar General Census Commissioner, India. Government of India Ministry of Home Affairs, n.d. Web. 20 June 2016. <http://www.censusindia.gov.in/2011census/dchb/Maharashtra.html>

High-Rise Organicity: A Proposal for a Bamboo-Themed Skyscraper

Xi Yue (Kelly)

Author background: Xi Yue (Kelly) grew up in China and currently attends The Affiliated High School of South China Normal University, located in Guangzhou, China. Her Pioneer seminar topic was in the field of architecture and titled "The Skyscraper."

1. Introduction

A strategy for an architect who wishes to catch people's eyes is to make it visually attractive, since it is generally appearance that determines one's first impression of an object. Some architects build remarkably tall structures that breakthrough the existing skyline in local cities with the aim of impressing people around the world and transmitting and heralding certain symbolic meanings required by their employers.

Because of people's "intense interest in reaching for the sky," a competition for height among skyscrapers has emerged.[11] The 828-meter tall Burj Khalifa Tower in Dubai, designed by Adrian D. Smith and completed in 2010, is a perfect example of such a structure, since it features a functionless, 250-meter-tall top, suggesting that the top is added to increase the aggregate height.[12] The Dubai government built this tall tower in order to "lure tourists and garner international interest."[13] Similarly, the Jeddah Tower (also known as the Kingdom Tower) in Jeddah, which is designed by Adrian D. Smith and still under construction, is even more ambitious in grabbing the title of "the world's tallest building:" its planned height is roughly 1000 meters.[14] Much taller than the currently tallest building Burj Khalifa Tower, the Jeddah Tower could trigger worldwide attention because it challenges both modern technology and human imagination: it is an easy task to construct something hundreds of meters up in the sky. The method to construct a building hundreds of meters up in the sky and the technology needed to keep the skyscraper safe have to be considered by engineers and architects.

[11] Ford, Larry, "The Diffusion of the Skyscraper As An Urban Symbol," *Yearbook of the Association of Pacific Coast Geographers* 35 (1973): 49-60.

[12] Council on Tall Building and Urban Habitat, "Burj Khalifa," *The Skyscraper Center*, at http://www.skyscrapercenter.com/building/burj-khalifa/3 [accessed 28 September 2016].

[13] Tingwei Zhang, "Importing Urban Giant: Re-Imaging Shanghai and Dubai with Skyscrapers," *International Journal of Architectural Research* (July 2013): 22-42.

[14] Peter Weismantle, and Alejandro Stochetti, "Kingdom Tower, Jeddah," *CTBUH Journal 2013 Issue I*, at http://www.ctbuh.org/TallBuildings/FeaturedTallBuildings/ArchiveJournal/KingdomTowerJeddah/tabid/4415/language/en-GB/Default.aspx [accessed 29 September 2016].

Figure 1. Adrian D. Smith, Burj Khalifa Tower, Dubai, 2010.

Figure 2. Adrian D. Smith, Jeddah Tower, Jeddah, unfinished

Some skyscrapers are symbols that reinforce cultural meaning. Their cultural value, either related to a company's public image or a turning point in architectural history at large, may cause people's curiosity to explore both the building and the message it conveys. In this case, the buildings are actually inviting passers-by to communicate with it. For instance, the "monumental simplicity" of the Seagram building, designed by the prominent German

architect Ludwig Mies van der Rohe in the mid-20th century, embodies its creator's "oft-repeated aphorisms that 'structure is spiritual' and 'less is more.'"[15] Its transparent curtain enables passers-by to get a clearer view of what is happening inside the structure, thereby breaks the boundary between the world outside the building and the one inside and creates an interaction. The Seagram Building is one of the predecessors of minimalism in architecture. Its historical impact and its assertive promotion of simplicity in the use of the glass curtain wall consolidate its status in the world. Even in the 21st century, the Seagram Building is regarded as a landmark in New York City.

Figure 3. Ludwig mies van der Rohe, Seagram Building, New York City, 1958.

The XXX building, whose design is proposed by me, is not like the Burj Khalifa and the Jeddah Tower which aim at pursuing incredible height like, but emphasizing the importance of cultural value just like the Seagram Building. The XXX building is planned to be located in Guangzhou, China, and financed by an advertising firm I would call XXX Company as its headquarter. With a moderate but still noticeable height compared to skyscrapers nearby, and a goal to construct an organic –an organic design is either a stimulation or a recreation of the great nature - and emotionally engaging building that would make people both inside and outside the building feel happy, the XXX challenges the current skyscraper style, and fulfills the XXX Company's expectations. Consequently, my design should win the competition.

2. Requirements of the XXX Company

XXX is a local Chinese advertisement company, whose headquarters are located in Guangzhou. Its ultimate goal is to always find a balance between innovation and tradition, as well as provocation and harmony. XXX believes that transmitting information efficiently while aesthetically is one of the major focuses of advertising. It prefers to find inspirations in nature since natural products are impeccably harmonious. For example, in order to advertise a camera, the XXX Company created a commercial in which the audience saw the

[15] New York University. "Seagram Building," *New York University Website*, at http://www.nyu.edu/classes/finearts/nyc/park/seagram.html [accessed 26 September 2016].

world through the eyes of different animals – including fish, dragonflies, and dogs – all captured by an instant-shot camera. The company consequently created the sense that though artificially made, the camera was highly connected to living creatures, thereby making it more emotionally appealing than a cold, grey machine.

In my assumption, the XXX Company has already earned some positive reputation within China, so it is planning to enhance its visibility and influence in the global advertisement market. It requires an office building that reflects the culture of its enterprise and stands out from the headquarters of their competitors, just as an outstanding advertisement that greatly impresses its audience. It also demands that the building offer a high-quality work experience.

3. Location

Composed of 99 floors, the 450-meter-tall building named XXX funded by the world-famous advertisement company, XXX is situated in Zhujiang New Town, the central business district in Guangzhou, China.

Figure 4. Map of XXX building.

On the northern side of the Pearl River lie several eminent structures, including the Guangdong Museum, the 440-meter-tall Guangzhou International Finance Center (IFC), the CTF Finance Center, which holds the laurels as the tallest building in Guangzhou, and the Guangzhou Opera House, which was designed by the famous British architect Zaha Hadid and whose construction finished in 2009. What's more, located at the southern side of the Pearl River stands a television tower named Canton Tower, opposite the cultural and financial building complexes.

Figure 5. Zaha Hadid, Guangzhou Opera House, Guangzhou, 2010.

Figure 6. Mark Hemel and Barbara Kuit, Canton Tower, Guangzhou, 2009.

There are two subway stations nearby; with an aggregate of eight entrances and exits in every direction, the transportation system around the XXX building is incredibly convenient. The closest entrance is only a 300-meter walk to the XXX building, which takes no more than five minutes. Since the XXX building is at the crossroads, people working in it have access to taxis as well. In addition to ample public transit, the underground parking lot holds space for private cars. The agency's employees' distinct demands for transport are therefore satisfied.

Beyond these pragmatic considerations lie two other reasons for choosing this location. First, the ingenious location of the XXX building triggers a dialogue between it and the Canton Tower. The dialogue which would be provoked by the two buildings' similar heights - the principal part of the Canton Tower rises to 454 meters, whereas the height of the XXX building is 450 meters - could create a sense of harmony and consistency. The two structures seem like two giants silently gazing at each other from each side of the river; although different in style, they share the same loneliness as unusual high-rises that are

much more conspicuous than any other constructions around them. The beautiful scene is more obvious at night, when viewers are able to perceive the two twinkling perpendicular lines from a great distance, recognizing them with no difficulty. The communication between the Canton Tower and the XXX building generates the interaction between the XXX building and its surroundings during day and night.

4**. Inspiration**

Due to their connection to advanced technology and their high production cost, skyscrapers almost always become the representation of prosperity, capitalism, and modernity.[16] Human beings sense their insignificance when standing in front of these cold colossal structures. There is little trace of life in most forms of high-rises other than the humans themselves, and nature is separated from the artificial world of skyscrapers in many cases, just like the Seagram building (see Fig. 3) and the Shanghai World Finance Center (see Fig. 7) which was built in 2008 in Shanghai and designed by Kohn Pedersen Fox Associates and Irie Miyake Architects & Engineers (IMA). Firstly, the skyscrapers are covered by glass and metal, two cold and bloodless materials that have no relationship with natural beings. Their metallic texture reflects a sense of rigidity and impassibility, further separating itself from human emotional world and organicity. Moreover, their strictly uniform outlook deprives them from defect and shows a highly artificial property. Through the two ways discussed above, nature is stripped away from these designs. The current state is undoubtedly unwholesome because nature is an indispensable part of human lives. It is time to reform.

Figure 7. Kohn Pedersen Fox Associates and Irie Miyake Architects and Engineers, Shanghai World Finance Center, Shanghai, 2008.

The idea of the XXX building is inspired by Guangzhou Opera House, which originated from coarse rocks and then underwent a process of recreation by the great

[16] Jean Gottman, "Why the Skyscraper?" *American Geographical Society* 56.2 (1966): 190-12.

architect, Zaha Hadid.[17] The exterior of the opera house is composed of a glass and steel frame that resembles a pair of enormous stones, which are humble yet precious treasures half-buried in the ground, opposite the high-profile skylines. This building shows that sometimes the best way to attract attention is not by featuring a flashy design, but by being organic and viewing the surroundings and the building itself as a unified entity. Hadid's masterpiece also suggests that the elements of sculpture can be applied to architecture which makes the structure more visually appealing. As a consequence, I have imitated the bamboo's shape in a bid to combine the vigor of the bamboo with the high-tech essence of skyscrapers, since they are both titans within their own species.

Bamboo is an important motif in traditional Chinese culture. It has such a practical value that it was applied to almost every area in ancient Chinese people's lives, products such as bamboo houses, bamboo baskets, bamboo hats, and bamboo clothes could be easily found. Bamboo was also a common ornamental plant in ancient Chinese landscape. Apart from its practicability, bamboo is also a representation of uprightness, tenacity, and tolerance according to Chinese culture. Whenever a Chinese sees a bamboo, he will naturally associate it with integrity. The same effect would be achieved when passers-by see the XXX building, a bamboo-shaped high-rise. As a result, since bamboo is deeply rooted in Chinese culture, it could help the XXX building automatically win a positive reputation.

I create an intriguing communication between the Guangzhou Opera House, the Pearl River, and the XXX building. They respectively resemble three significant motifs commonly found in ancient Chinese painting: stone, water, and bamboo. Thus, they reinterpret the traditional value of art. This use of traditional elements in modern structures gives new meanings to the classic Chinese landscape. Such implicit coordination could arouse the audience's infinite curiosity and imagination and make it suddenly realize that the XXX building is more than an office building, and that the connotations of the Opera House are waiting to be discovered. The skyscraper's effective interactions with the environment, with both artificial structures and rivers, are meant to build an undivided and organic whole.

The Taipei 101 building in Taiwan, which was designed by C.Y.Lee & Partners in 2004, acts as a model for mingling Chinese traditional motifs with a form of architecture originated in Western countries. It reminds people that we should always keep our nation's traditions in mind while interacting with foreign cultures. It requires much effort since Chinese traditional buildings usually expand horizontally while the high-rise extends vertically. But Taipei 101 cleverly picks out the most typical notions such as the Chinese lucky number 8, the "ruyi" figure which represents fulfillment and pleasure, and resemblance to the design of ancient Chinese towers. Learning from it, I incorporate some reference to a past architectural style when designing the southern door of the XXX building.

[17] Sanjay Gangal, "Guangzhou Opera House in China by Zaha Hadid Architects", *AECCafe Blogs*, at http://www10.aeccafe.com/blogs/arch-showcase/2011/03/03/guangzhou-opera-house-in-china-by-zaha-hadid-architects/.

Figure 8. Zuyuan Li and Chongping Wang, Taipei 101, Taipei, 2004.

The Mori Art Museum on the 53th floor in the Mori Tower is a contemporary art museum in Tokyo, Japan.[18] It is a foreign counterpart to the art museum in the XXX building, for they are both "art museum in the sky." Though they both exhibit contemporary artworks and share a similar height, the art museum in the XXX building is even better since visitors to the Mori Art Museum are isolated from outside scenery, which weakens its advantage in height; contrarily, in the art museum in the XXX building, there are six windows abreast of the paintings and photographs. A perfect combination of artworks and outside view is thereby reached. The scarcity of formal art museum in Guangzhou also inspired me to construct an art museum which could fulfill Guangzhou citizens' demand for a broader variety of art experience.

5. Basic Information

The XXX building is a 450-meter-high cylinder, whose diameter is 100 meters. I apply a relatively new technology named tube-frame construction to its entire structure, which creates larger room for interior space while reducing lateral load brought by wind.[19] The skyscraper has 99 floors, and it is divided into three sections: the base (B5 - 2F), the middle (3 - 97F), and the top (98 - 99F). The base includes the lobby on the first floor, a morning tea restaurant on the second floor, an underground shopping center on B1, and an underground parking lot from B5 to B2. What's more, a luxurious French restaurant would be located on 98F at the top, and an art museum would assume the dominant position on 99 F. The rest of the stories would be rented out as offices except 96 - 97F, which would be used by the XXX Company itself.

6. Exterior Description

The XXX building is in the shape of a bamboo being cut at an angle in the middle. From its exterior, viewers could see the divisions of the skyscraper: the base, the middle, and the top. The division roughly correlates to the different functions of the skyscraper: the middle is for commercial use, whereas the top and the base are for consumption.

[18] For more information on Mori Art Museum, see http://www.mori.art.museum/eng/index.html.

[19] Skidmore, Owings & Merrill LLP's. "DEWITT CHESTNUT APARTMENTS," SOM.com, at http://www.som.com/projects/dewitt_chestnut_apartments.

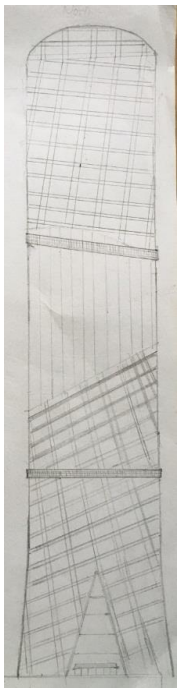


Figure 9. Northern fa çade of XXX building.

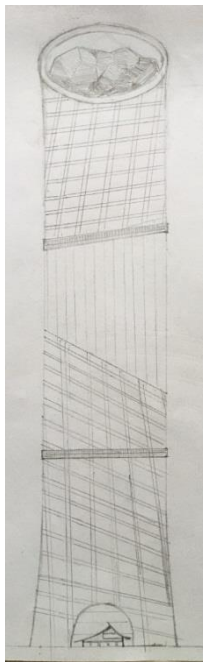


Figure 10. Southern fa çade of XXX building.

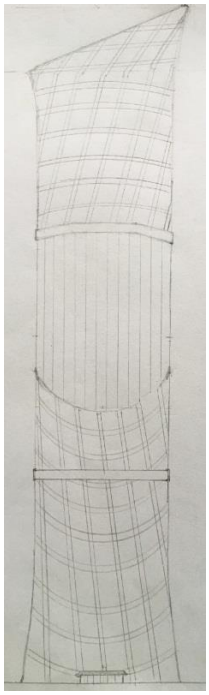


Figure 11. Eastern fa ̧ade of XXX building.

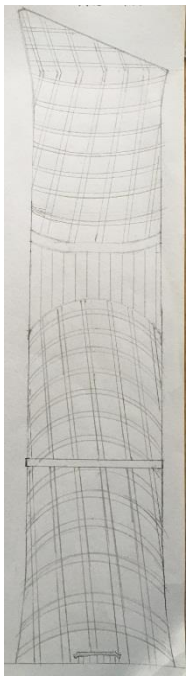


Figure 12. Western fa ̧ade of XXX building.

The design of the XXX building consists of diagonal lines. The base and top of the skyscraper are made up of vertical columns as well as horizontal beams, which seem like lines when viewing the structure as a whole. The vertical lines connecting the top to the bottom are generally thicker than beams. Their lower parts are even thicker than the upper, which gives off the feeling that these bright green bars are validly rooted in the ground. A sense of loftiness is created through the columns as well. The horizontal lines are parallel to each other to better set off the tilt vertical lines. Behind the concrete bars is a cylindrical glass curtain wall. At night, the LEDs stuck on those concrete bars would emit luminously green light and the bright yellow light inside the structure would be seen through the transparent glass. Due to the light effect, the XXX building would look like a giant bright yellow bamboo with green lines on its surface.

To avoid visually competing with the base, the middle section of the skyscraper is no more intricate than some long, wood-colored iron panels pieced together. Square windows are embedded in these panels following the same sequence. The width of each plank is one meter wider than a window, so there is a one-meter interval between each two adjoining windows.

Beyond supporting the whole building, the tube-framed structure is also a component of the bamboo. Three cold-grey spandrel beams represent bamboo joints: two are set at the point of transition between the top and the middle, and between the middle and the bottom; the remaining one is located approximately two-thirds into the bottom section. Except for the lowest one which is parallel to the horizontal plane, the other two beams are either downward sloping to the right or inclining to the left. The asymmetrical arrangement is as organic as those natural creations, thereby contributing to the organicity of my design. Despite the fact that these asymmetrical lines go against Chinese traditional architecture's preference of symmetry, they help weaken humans' intervention, and make the XXX building more like a naturally created existence.

The exterior design of the top of the XXX building is similar to the bottom, so that the two establish a sense of uniformity between the root and the pinnacle, enforcing a sense of completeness. The apex of the high-rise is hollow, and its edge is slightly flaring like a blossoming flower (see Fig. 6 and 7). The open mouth of the bamboo has a shape like a water-drop, and its tapered end faces the south, reflecting that the XXX building is willing to communicate with the Canton Tower on the south. Beyond that, south-facing building is generally preferred by people in southern China since it correlates to a higher exposure of sunlight and wind. As a result, opening the mouth to the south symbolizes that the XXX Company as well as other organizations in the building enjoy a bright future.

Another unique feature of the XXX building is that it houses a spherical art museum on the 53rd floor surrounded by floor-to-ceiling curtain wall windows. All three sides of the art museum are open to the outside, except the northern side where it is wrapped in the tapered end of the water-drop. Together it looks like a piece of grey stone placed in the mouth of a bamboo. I shape the art museum into a dark stone so that it could turn into a counterpart of the Guangzhou Opera House, though much smaller in size. It could better blend into the local environment, which again corresponds to the concept of harmony. Apart from that, considering that diversity is a typical feature of the nature, I add the stone to the top of the bamboo to increase the sense of heterogeneity of the whole building: it is not a lonely plant in the city, but a diverse community including stones and a bamboo. Thus the XXX building becomes more organic and its existence becomes more reasonable and natural.

Moreover, I incorporate Chinese elements into the southern door (see Fig.7), the one facing the Pearl River and opened to the customers. Firstly, the large arch on the wall

represents a traditional Chinese door opening, frequently found on doors of palaces and aristocrats' houses. What's more, the cornices of the door are in the shape of a Chinese character, "入," the meaning of which is "to enter," correlating to the function of the entrance door. There is a long window under the Chinese character, through which visitors on the second floor could see the outside world. Apart from that, there is an entrance portico comprised of six columns door. The columns resemble bamboo sticks, which also remind passers-by that they are next to the famous bamboo-like skyscraper, the XXX building. The other three doors are all alike in their appearance - similar glass doors, and a traditional Chinese lintel over them. They are different from the southern one for economic reasons.

7. Interior Description

There are 20 elevators and two escalators inside the building. Each of the elevator is only available to certain floors to ensure its own efficiency during rush hours. The two escalators and four of the 18 lifts are for customers of the shopping center, two restaurants and the art museum; the remaining are "staff-only," reserved exclusively for people who work in the skyscraper. For security reasons, there are eight egress staircases: two on each side, next to each gate.

A. Lobby

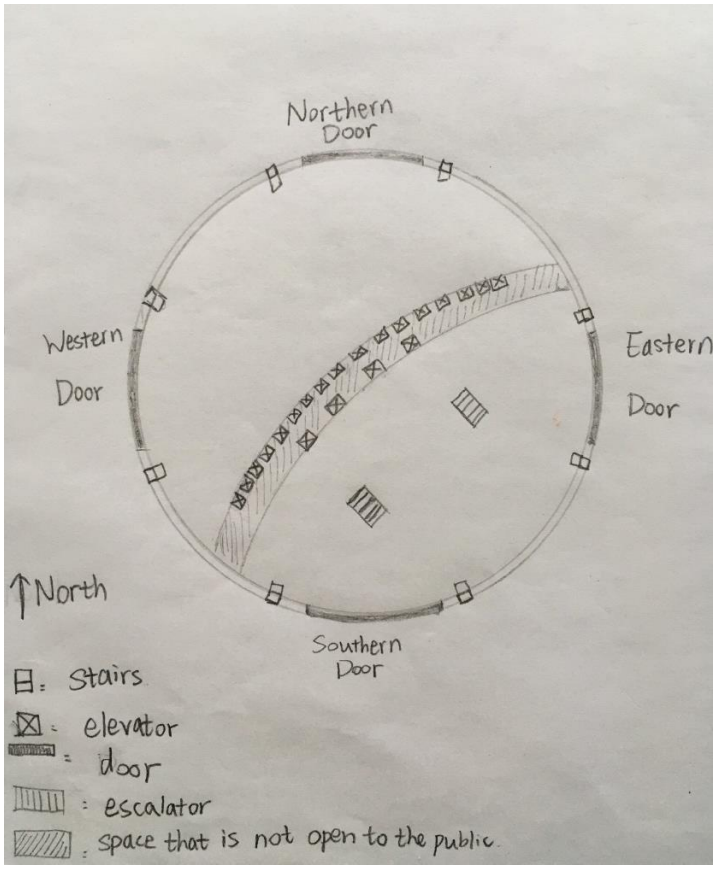


Figure 13. Ground plan of the lobby of the XXX building.

The lobby contains four gates, and is separated into two distinct areas by a curving wall concaved towards the south-east direction in the middle of the first floor. The side facing the Pearl River, which contains the north gate and the east gate, is for customers of the shopping center, restaurants, or the art museum; the other half is the lobby for commercial use. The 16 elevators are embedded in the curved wall, as shown in the above ground plan.

B. Morning Tea Restaurant

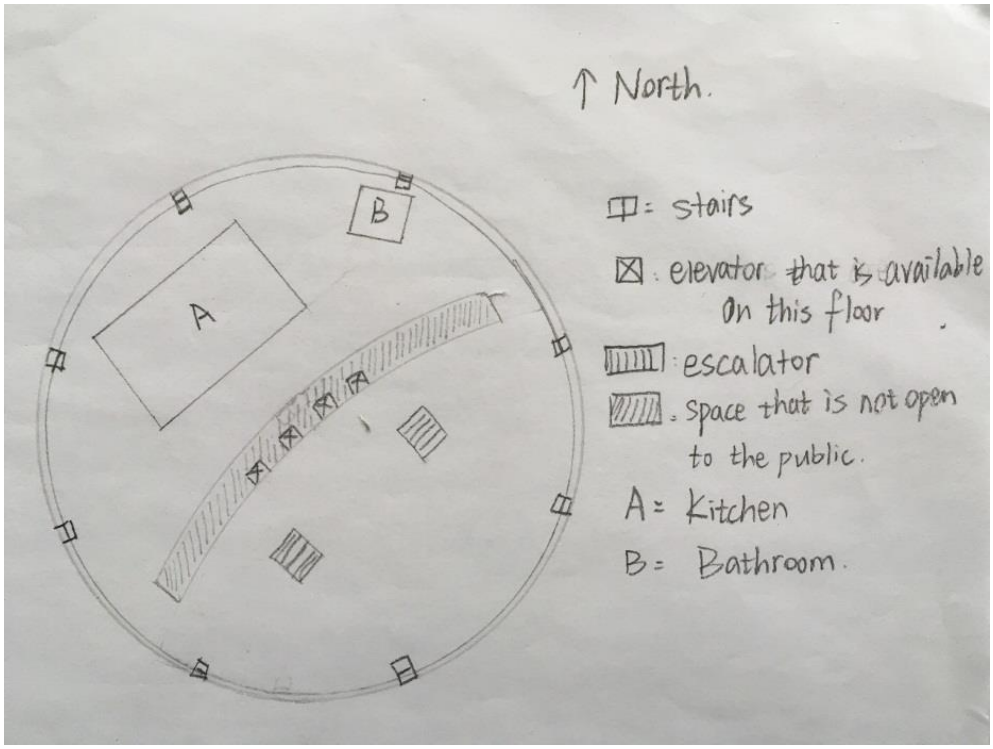


Figure 14. Ground plan of the morning tea restaurant of the XXX building.

As shown in the ground plan above, the elevator shaft is no longer connected to the interior wall of the building. I remove the two parts to provide more space for dining tables and walking. A rectangular restroom is set on the convex side of the curving wall. The kitchen occupies a rectangular room in the northern part. The rest of the space is left for dining tables, chairs, and passageways.

C. Offices

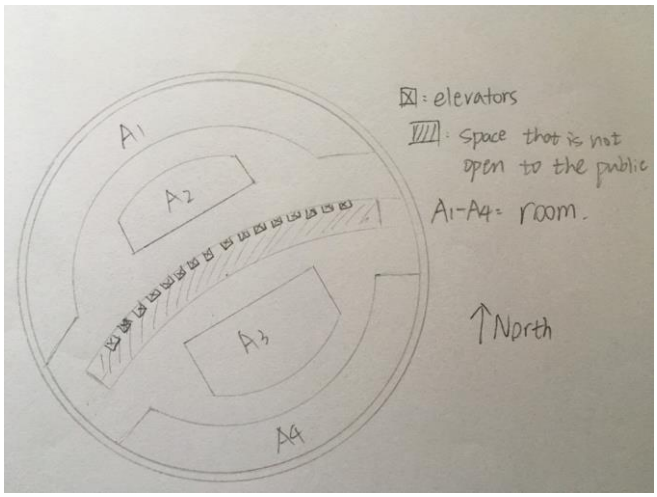


Figure 15. Ground plan of the offices of the XXX building.

The offices on the following floors are unique in that their edges are curved along the windows and wrapped around the inner core, which houses the elevator shaft. Each floor is generally divided into four rooms and a public bathroom near the elevator.

D. French Restaurant

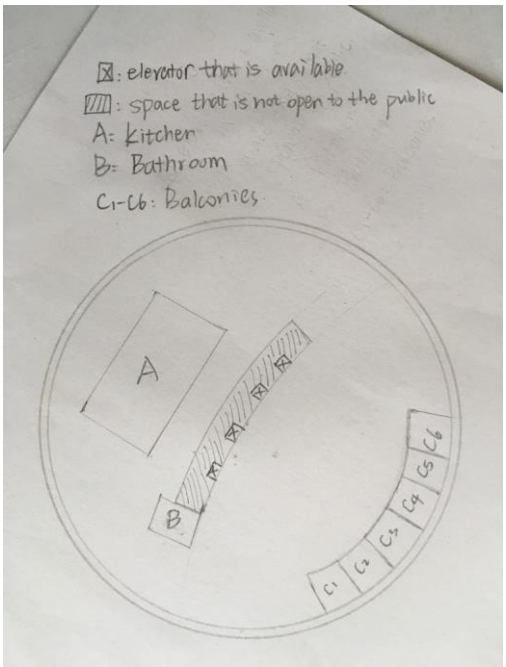


Figure 16. Ground plan of the French restaurant of the XXX building.

The French restaurant at the top has a similar layout to that of the morning tea restaurant in terms of the location of the kitchen, the bathroom, and the elevators, since they are alike in function. By doing so, the harmony of spatial distribution is preserved. The wall is completely made of glass, which makes it seem like a transparent box. Customers are thus capable of having a 360 ° view accompanied by the absence of blind corners. Six balconies open up to the south, facing the Pearl River. Though alike in function, the French restaurant and the morning tea restaurant target different clients. The wealthy people are the major clients of the French restaurant due to its expensive food and excellent position, whereas average Guangzhou citizens could afford to eat in the morning tea restaurant.

E. Art Museum

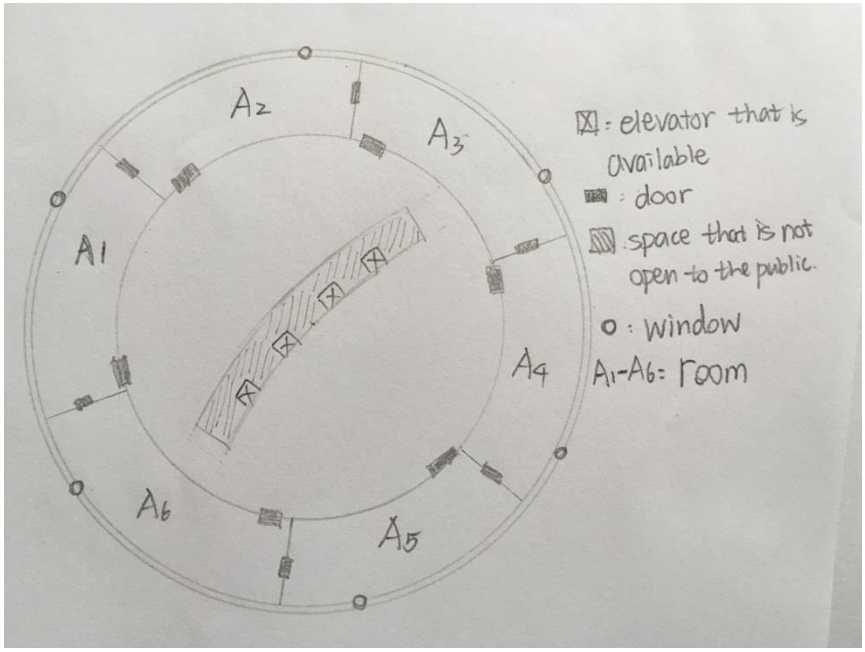


Figure 17. Ground plan of the art museum of the XXX building.

There are six rooms in the art museum, and each of them is connected to the neighboring two by doors. I try to reduce the interaction between the audience and the outside world when designing the art museum or it is probable that people would be distracted by the rarely seen bird's-eye view and stop enjoying the artworks. In order to seize the unique opportunity provided by its location and prevent the potential competition between artworks and outside scenery, the gallery would feature a total of six windows installed parallel to the contemporary artworks. The visitors are not able to see the outside environment except through the six windows, thereby reducing the possible distractions audience may get. I do not delete all windows because the art museum's advantage in height would otherwise be wasted. These windows would be shaped differently in a bid to best exhibit the beautiful scenery outside the thick walls without destroying the harmoniously artistic atmosphere.

8. Impact

The XXX building could fulfill all of the requirements of the XXX Company, and it could also go far beyond that.

First, the morning tea restaurant could reduce the absolute separation between office and home, so working in this skyscraper would be less alienating, and feel more accessible and humane. Before the Industrial Revolution, most people worked at home, which enabled them to frequently communicate with family members. After industrialization, however, the working place shifted from homes to large factories where workers gathered and did their jobs. The morning tea restaurant is set to decrease this phenomenon that interaction between family members was reduced as a result of the separation of home and working place. According to the local Guangzhou tradition, drinking morning tea is often related to family gathering. Families will often go to a morning tea restaurant on weekends and sit together, chatting about recent news. Nevertheless, on weekdays, only older, retired people could engage in this practice. So by introducing the Guangzhou Restaurant, the most popular morning tea restaurant in Guangzhou, to the XXX building, I could offer an ideal area for the employees' parents and grandparents to chat with each other and taste delicious Canton delicacies while the younger generations are working a few floors above.

The connection that ties the XXX building with its surrounding structures creates harmony and consistency. A new interpretation of the Canton tower would emerge when the XXX building is constructed. Because the heights of the two structures are so close, they seem to compete as well as support each other at the same time ---- a more human message than "a TV tower that radiates and receives signals" or "a tower which resembles a thin waist." The romantic relationship between the Pearl River, the Guangzhou Opera House, and the XXX building legitimizes its existence and makes citizens accept such a futuristic construction.

The XXX building could stand out from nearby skyscrapers since it is the only organic one in that region. Its uncommon sculptural appearance inspired by the bamboo plant distinguishes it from other similar and conventional skyscrapers, the design of which fails to strive for citizens' emotion. Then, naturally, as more and more people become aware of its unique appearance, as awareness of this skyscraper spreads, the company would receive more attention. At that moment, not only the XXX Company would receive publicity, but tourism in Guangzhou would be rapidly boosted, attracting many visitors.

Once the XXX building is known to foreigners, it may cause them to develop a new perspective on Chinese traditions, and show them that Chinese culture is not as conservative as many people still believe, and that it does not stand in opposition to contemporary architecture. A "two-way communication of ideas and energies" between China and the west is therefore established, that China could gain a status in the study of modern architecture.[20]

9**. Conclusion**

In conclusion, the XXX building could fulfill the company's requirements for creating a good working environment, harmonious design, and unique outlook. It could promote tourism in Guangzhou and help spread an understanding of Chinese traditional culture.

[20] Jianfei Zhu. "Criticality in between China and the West," *The Journal of Architecture* 10:5 (2005): 479-498.

Su Shi, a well-known Chinese poet, wrote, “People would rather eat without meat than live without bamboo.” The XXX building, a modern display of an ancient motif, is equally necessary to modern Guangzhou, China, and the world at large.

Bibliography

Council on Tall Building and Urban Habitat, "Burj Khalifa," *The Skyscraper Center*, at http://www.skyscrapercenter.com/building/burj-khalifa/3 [accessed 28 August 2016].

Ford, Larry, "The Diffusion of the Skyscraper As An Urban Symbol," *Yearbook of the Association of Pacific Coast Geographers* 35 (1973): 49-60. Web.

Jean Gottman, "Why the Skyscraper?" *American Geographical Society* 56.2 (1966): 190-12.

Jianfei Zhu. "Criticality in between China and the West," *The Journal of Architecture* 10:5 (2005): 479-498.

New York University. "Seagram Building," New York University Website, at http://www.nyu.edu/classes/finearts/nyc/park/seagram.html [accessed 26 September 2016].

Peter Weismantle, and Alejandro Stochetti, "Kingdom Tower, Jeddah," *CTBUH Journal 2013 Issue I*, at http://www.ctbuh.org/TallBuildings/FeaturedTallBuildings/ArchiveJournal/KingdomTowerJeddah/tabid/4415/language/en-GB/Default.aspx [accessed 29 September 2016].

Sanjay Gangal, "Guangzhou Opera House in China by Zaha Hadid Architects", *AECCafe Blogs*, at http://www10.aeccafe.com/blogs/arch-showcase/2011/03/03/guangzhou-opera-house-in-china-by-zaha-hadid-architects/.

Skidmore, Owings & Merrill LLP's. "DEWITT CHESTNUT APARTMENTS," *Skidmore, Owings & Merrill LLP's*, at http://www.som.com/projects/dewitt_chestnut_apartments.

Tingwei Zhang, "Imorting Urban Giant: Re-Imaging Shanghai and Dubai with Skyscrapers," *International journal of Architectural Research* (July 2013): 22-42.

Molecular Transistors: The Effects of Test Molecules on the Conduction Patterns of a Lithium Nanowire

Isaac Ick

Author background: Isaac Ick grew up in the United States and currently attends Dobyns-Bennett High School, located in Kingsport, Tennessee. His Pioneer seminar topic was in the field of chemistry and titled "Computational Quantum Chemistry."

Abstract

The conductivity patterns of certain molecular materials are a highly studied part of quantum mechanics. The behavior of electric current as it is conducted through a material can easily change based on applied voltage. This paper focuses on the effects of different test molecules on the conductivity of a lithium nanowire. When different test molecules are attached to the lithium nanowire in computer simulations, varying voltages applied on the apparatus stimulate different current outputs. This behavior could indicate that a new type of molecular semiconductor device has been observed.

Introduction

Semiconductors are basic devices heavily utilized in modern electronics due to their unique band gap energy properties. Band gap energy refers to the necessary voltage applied to a material in order to make that material conductive. On a molecular level, the band gap energy is defined as the amount of energy required to move an electron from the valence band to the conduction band. In insulating materials, the band gap energy is very large. Even with a high voltage, insulators refuse to conduct electricity. On the other hand, conducting materials have band gap energies of zero; they require very little voltage and do not require an applied voltage to conduct electricity[1]. In addition to this, semiconductors are described as materials having small band gap energies, typically between 0 and 4 electron volts[2].

Materials possessing this quality are often made of silicon; however germanium is an ample substitute. There are three main types of semiconductor devices made of silicon. They can be derived from the relationship between applied voltage and produced current. The first type is known as the transistor. Most transistors display a linear relationship between voltage and current. As voltage increases, so does current. However, once a certain voltage threshold is reached, transistors cease to conduct more current. In most electronics, binary information is stored using transistors. When in its resistive state, the transistor is in the 0 position. When voltage is applied and it reaches its conductive state, the transistor is in the 1 position[3].

Diodes are similar semiconductor devices. They do not conduct electricity until a certain voltage is reached. However, diodes can conduct electricity easily in the forward direction. After the breakdown voltage is reached, diodes become very conductive in the reverse direction. Diodes are utilized in the form of LED's, photodiodes, signal diodes, and vacuum diodes and there are many other types[4].

A third application of semiconductor that is very useful to modern society is the switch. A switch has a conductive state and resistive state based on different external variables. Some

switches respond to physical or chemical stimuli. When the switch is stimulated, it will change states from conductive to resistive or vice versa. Altering a switch can change the efficiency with which the switches operate. For example, switches could change states at different voltages and at different speeds[5].

Finding molecules that display transistor-like qualities is the goal of this investigation. Transistors are important semiconductor devices because they can store binary information through their conductive and resistive states[3].

When first utilized, transistors were much larger. Therefore, larger computers were needed to contain these larger transistors to perform calculations. However, as technology has progressed, transistors have become smaller and smaller. This had led to computers that continue to decrease in size while utilizing an ever-increasing number of transistors. With a larger number of transistors, computers store more information and perform operations faster[6].

An issue with today's semiconductor devices is that they are becoming too small. Quantum effects are starting to interfere with the function of tiny transistor devices[7]. As transistors have become smaller, the distances electrons travel from emitter to collector have also decreased[6]. When the Schrödinger equation is solved for a particle approaching a barrier, smaller transistors reveal an issue:

$$\frac{-\hbar^2}{2m}\frac{\alpha^2\psi(x)}{\alpha x^2} = (E - U_0)\psi(x)$$

$$\psi = Ae^{-\alpha x} \; where \; \alpha = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}}$$

Here, U_0 is a variable that describes the barrier's height and thickness. Therefore, as the height and thickness of a barrier decreases, the probability of tunneling increases[8]. If all of the regions between the emitter and collector of a bipolar junction transistor are considered a barrier, then by shrinking the transistor, the probability of tunneling increases. This means that in modern electronics, more and more electrons are tunneling through transistor devices, creating false currents. When such a current is detected, the transistor enters its conductive state, and records incorrect binary information. For this reason, smaller transistors cannot be made, as they will no longer function[9].

This becomes an issue because faster computational times are required for further advancing fields of science. However, if transistors remain the same size, computers will have to become larger. In order to reduce the size of electronics devices, a new generation of transistors and other semiconductor devices must be discovered. The goal of this project is to possibly discover a smaller, feasible transistor to replace the larger ones.

A new generation of transistors could prove completely unique from the original devices in their current output, or they could be relatively similar. It is possible that the new type of transistors will follow the traditional patterns and will be easily recognized by its current graph. However, it is also possible that the computer created device will simply reduce the band gap energy necessary to operate a transistor device. Band gap energy in transistors can be described as the voltage required to change the transistor from resistive to conductive as a whole apparatus. However, it can also mean the energy required to move an electron from the valence band to the conduction band (see Figure 1).

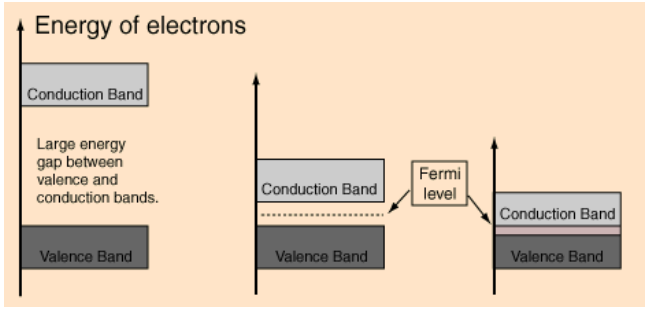


Figure 2: *Band gap energy based on valence and conduction band energies. The vertical axis represents energy.*[10]

Much research has been done to artificially alter the band gap energy of transistors and semiconductors as a whole. Reducing band gap energy reduces the amount of energy required to operate many modern electronic devices. Therefore, by lowering the band gap energy, there is a potential to reduce electricity consumption by a massive amount and on a global scale[11].

A popular topic has been to study the effects of temperature on the band gap energy of different semiconductor devices. In general, as temperature increases, the band gap energy decreases. This is due to weakening bonds allowing electrons to be moved to the conduction band with less energy[12]. These studies are relevant to this investigation because they minimize the number of variables that need to be tested. Different conditions like temperature, pressure, and structural variables should remain constant when testing for band gap energy in order to ensure the results are known only for those conditions.

It is likely that a new type of transistor can be created on the molecular level. Quantum principles will likely cause strange current outputs when different voltages are applied. Some of these seemingly strange principles, however, could mimic a traditional semiconductor device and be of use. However, the feasibility of implementing these devices is slim. Right now, the semiconductor devices that are being manufactured are extremely small: transistors often measure only 10 nanometers[13]. While the manufacturing industry has been comfortably engineering tiny devices, the new generation of transistors and semiconductors will only be a couple of atoms large. It is not possible to manufacture transistors of this size yet, so creating a single device designed in this investigation would be extremely difficult.

Methods and Materials

In the beginning of this investigation, three different test molecules were chosen: trichloroethylene, benzene, and aniline. There was no particular emphasis on choosing certain types of molecules. A wide variety of properties, characteristics, and atom types were represented in this group, making it a well-rounded choice.

First, the band gap energy of these molecules was tested using Gabedit[14] as a graphical interface and NWChem[15] as the computational program. Gabedit was opened first, and the desired molecule was drawn on the drawing board. Afterwards, the drawing was formatted into a run file for NWChem through the use of Cartesian coordinates and different density functionals. The functional methods used were Hartree-Fock, BYLP, and B3LYP. The purpose of each functional was to optimize the structure of the molecule in order to produce the most accurate band gap energy. However, some functionals work better than others. Each functional has a different way of calculating various vectors and energy values.

The biggest difference between each functional method is its ability to deal with electron correlation[16]. Electron correlation is a part of how electrons interact when in close proximity in quantum systems[17]. Electron correlation is difficult to calculate with high precision, especially in a system with many electrons. Therefore, functional methods that calculate band gap very precisely and with high attention to electron correlation take a very long time. The functionals chosen in this investigation were picked due to manageable computational times paired with accuracy.

After the run file was made, it was sent to NWChem to run calculations. NWChem produces output files that organize orbital energies into vectors. Each vector has a calculated energy value that corresponds to the energy of electrons within certain molecular orbitals. The difference in energy between the vectors that represented the valence band and conduction band gave the band gap energy. Most materials, however, contain electrons up to the highest valence energy[18]. Therefore, the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) was determined to be the band gap energy. The band gap was measured when different voltages were applied to each molecule.

Using SIESTA[19] and tranSIESTA[20], a lithium nanowire was optimized. The wire was only ten atoms long and only one atom thick. Four atoms on both ends of the wire were optimized to become the electrodes, where the current enters the left and exits on the right side of the system[21]. The .TSHS output file from the electrode calculation was needed for the voltage testing in order for each side of the wire to act as an electrode.

Before voltage testing, however, the lithium wire had to be optimized in real space too. Therefore, calculations took place to determine the optimal lattice constant and mesh cutoff for the lithium wire. The lattice constant refers to how far away each atom is spaced from the next atom and is useful in structures with precise spatial measurements[22]. When the optimal lattice constant was reached, an extremum in total energy was observed. In this case, the optimal lattice constant value for the apparatus was 3.0 angstroms. This optimizes the structure of the apparatus to a lower total energy, making current through the designed apparatus more realistic. Mesh cutoff is a measurement that adjusts real space in order to enhance structural characteristics of molecules[23]. Again, the purpose of this optimization is to make sure that the lithium wire is structurally apt to conduct current in a realistic manner. The optimal mesh cutoff is also determined by an extremum in total energy. After finding the optimal lattice constant and mesh cutoff, the electrode calculation was run again in order to perfect the electrode structure and function. Then, current through the lithium wire was measured while varying the voltage from zero electron volts to two electron volts.

Next, individual test molecules were added to the lithium wire. In order to add a molecule to the lithium wire, the actual molecule had to be optimized in Gabedit with a lithium atom replacing one of the hydrogen atoms in the molecule. The lithium atom would act as a transfer point in order to recreate the molecule in SIESTA and tranSIESTA. After the structure was optimized, the Cartesian coordinates were obtained. The Gabedit coordinates had to be scaled considering Lattice Constants within the SIESTA and tranSIESTA run files. Also, in tranSIESTA, current is applied through the z direction, so a 90-degree rotation had to occur. The process for transferring coordinates is described below.

Table I. *Naming different atom coordinates in order to demonstrate a coordinate transfer.*

Atom	X	Y	Z
$Atom_1$	X_1	Y_1	Z_1
$Atom_2$	X_2	Y_2	Z_2
$Atom_3$	X_3	Y_3	Z_3
Lithium	X_{Li}	Y_{Li}	z_{Li}

SIESTA and tranSIESTA Transformation-

$$X_{Li} + X_0 = 0.3125 \times L_c$$
$$Y_{Li} + Y_0 = 0.3125 \times L_c$$
$$Z_{Li} + Z_0 = 0.4 \times L_c$$

where L_c is the system lattice constant
Solve for (X_0, Y_0, Z_0) to utilize in transfer.

Table II. *Equations for transferring Gabedit output coordinates into SIESTA and tranSIESTA input coordinates.*

Atom	X	Y	Z
$Atom_1$	$\frac{X_1 + Z_0}{L_c}$	$\frac{Y_1 + Y_0}{L_c}$	$\frac{Z_1 + X_0}{L_c}$
$Atom_2$	$\frac{X_2 + Z_0}{L_c}$	$\frac{Y_2 + Y_0}{L_c}$	$\frac{Z_2 + X_0}{L_c}$
$Atom_3$	$\frac{X_3 + Z_0}{L_c}$	$\frac{Y_3 + Y_0}{L_c}$	$\frac{Z_3 + X_0}{L_c}$
Lithium	$\frac{X_{Li} + Z_0}{L_c}$	$\frac{Y_{Li} + Y_0}{L_c}$	$\frac{Z_{Li} + X_0}{L_c}$

After the coordinates were added to the input files, a test run was conducted at zero voltage. The tranSIESTA executable gave an XV file which was converted into an .XSF file in order to be viewed using a program called xcrysden[25]. When the lithium wire is observed in xcrysden, the test molecule should be bonded to one of the middle two lithium atoms to ensure it has not become a part of the electrode. Correct bonding between atoms was verified. Oftentimes, however, xcrysden fails to create bonds between atoms with the same or similar z coordinates. This was the case with the trichloroethylene molecule. Even though a bond did not appear in the graphical interface, voltage testing was done because SIESTA and tranSIESTA treated the two atoms as bonded.

After the test molecule was attached to the lithium wire successfully, the current through the apparatus was measured at different voltages. Voltages started at 0 electron volts and increased in increments of 0.1 electron volts until 1 electron volt was reached. Then, the increment increased to 0.2 electron volts until 2.0 electron volts was reached. This method of slowly increasing the voltage helped to reduce computational time since SIESTA and tranSIESTA can draw information from similar files. After voltage testing was complete, all the necessary data was collected.

Data

Aniline Band Gap Energy-

Table III. *Electron energies of an aniline molecule based on charge and calculated using the B3YLP functional.*

Charge (eV)	Energy (HOMO)	Energy (LUMO)	Band Gap Energy
0	-55.363 eV	33.447 eV	88.809 eV
4.019	-55.396 eV	49.123 eV	104.519 eV
8.038	-55.628 eV	83.560 eV	139.188 eV
12.057	-55.783 eV	116.038 eV	171.820 eV
16.078	-55.928 eV	154.987 eV	210.915 eV
20.100	-56.083 eV	186.581 eV	242.663 eV
24.124	-56.225 eV	222.855 eV	279.080 eV
28.145	-56.371 eV	256.084 eV	312.455 eV
32.170	-56.511 eV	29.149 eV	85.660 eV
36.194	-56.649 eV	32.789 eV	89.438 eV

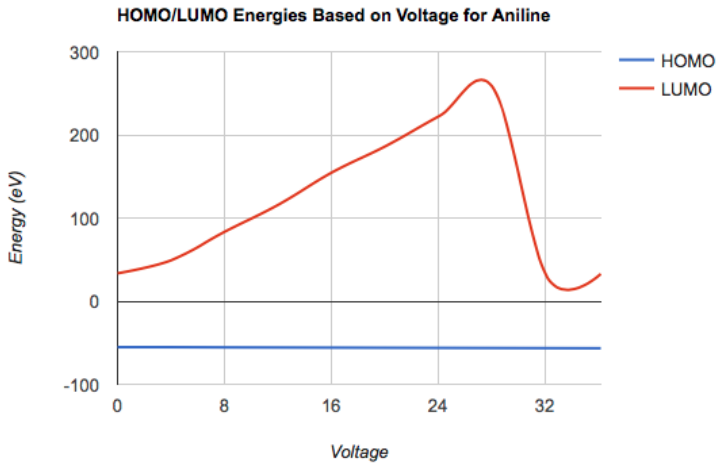


Figure 2: *Changing HUMO and LUMO energy values of aniline with applied voltage.*

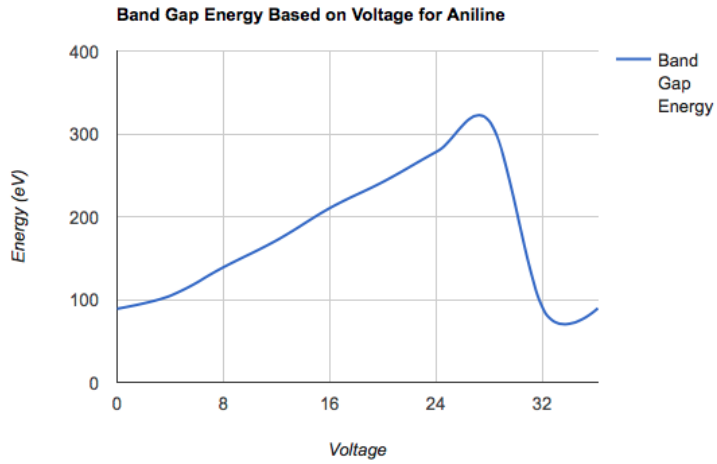


***Figure 3**: Changing band gap energy of aniline molecule with applied voltage.*

Benzene Band Gap Energy-

***Table IV.** Electron energies of a benzene molecule based on charge and calculated using the B3YLP functional.*

Charge (eV)	**Energy (HOMO)**	**Energy (LUMO)**	**Band Gap Energy**
0	-81.335 eV	-58.853 eV	22.482 eV
0.631	-179.725 eV	-68.863 eV	110.863 eV
1.888	-180.389 eV	-68.764 eV	111.626 eV
2.520	-180.560 eV	-68.716	111.844 eV
3.151	-180.630 eV	-68.670 eV	111.960 eV
5.050	-179.318 eV	-68.550 eV	110.768 eV
5.880	-176.391 eV	-68.494 eV	107.897 eV
7.602	-171.065 eV	-68.453 eV	102.612 eV
8.239	-169.106 eV	-68.445 eV	100.660 eV
9.513	-153.380 eV	-68.42 eV	84.956 eV

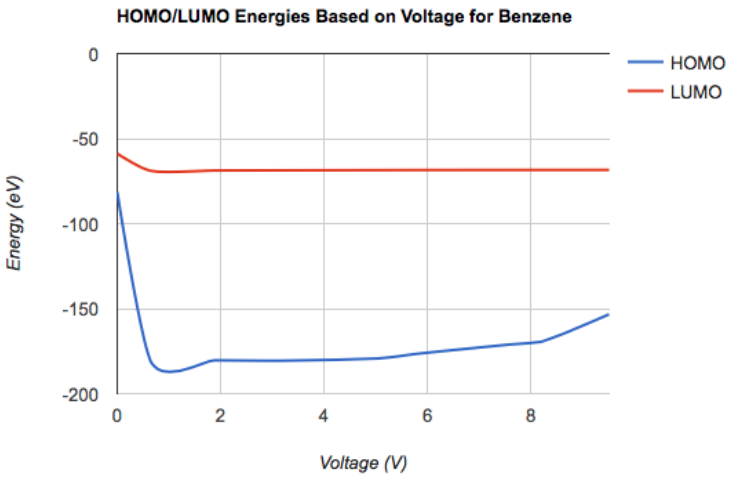


Figure 4: *Changing HUMO and LUMO energy values of benzene with applied voltage.*

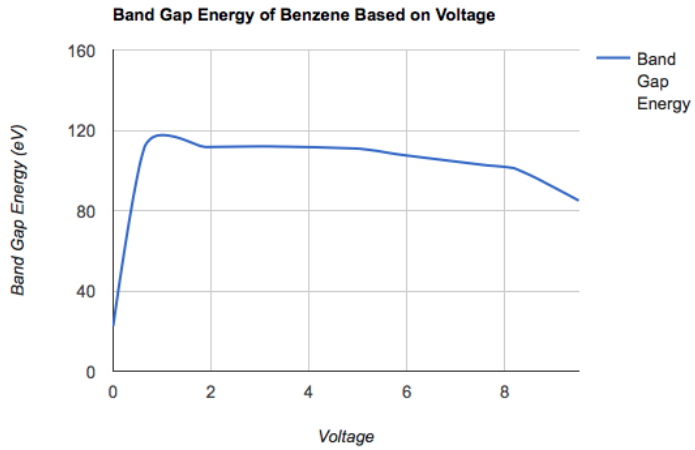


Figure 5*: Changing band gap energy of benzzene molecule with applied voltage.*

Trichloroethylene Band Gap Energy-

Table V. *Electron energies of a trichloroethylene molecule based on charge and calculated using the B3YLP functional.*

Charge (eV)	Energy (HOMO)	Energy (LUMO)	Band Gap Energy
0	-86.524 eV	-76.815 eV	15.709 eV
0.294	-86.624 eV	-70.819 eV	15.805 eV
1.472	-86.923 eV	-70.833 eV	16.090 eV
2.941	-87.302 eV	-70.851 eV	16.450 eV
4.408	-87.648 eV	-70.870 eV	16.778 eV
5.872	-88.012 eV	-70.888 eV	17.125 eV
7.336	-88.377 eV	-70.905 eV	17.472 eV
8.800	-88.731 eV	-70.923 eV	17.808 eV
10.261	-89.203 eV	-70.941 eV	18.263 eV
11.716	-89.470 eV	-70.957 eV	18.513 eV

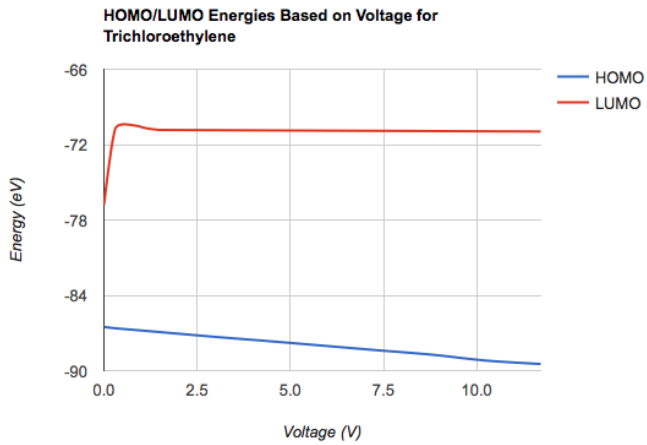


Figure 6: *Changing HUMO and LUMO energy values of trichloroethylene with applied voltage.*

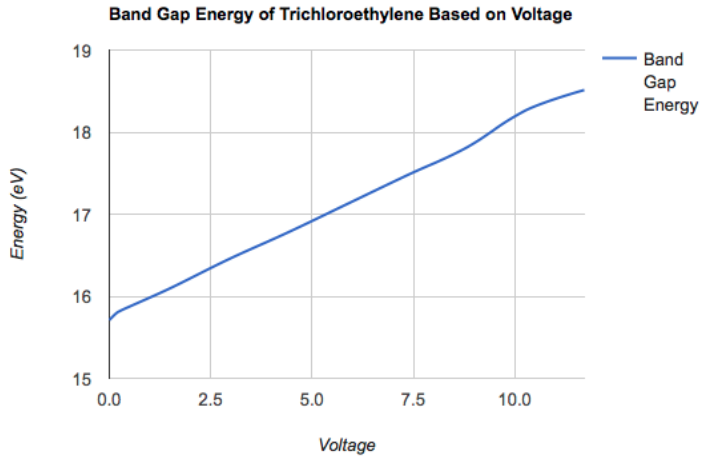


Figure 7: *Changing band gap energy of trichloroethylene molecule with applied voltage.*

Lithium Wire Optimization-

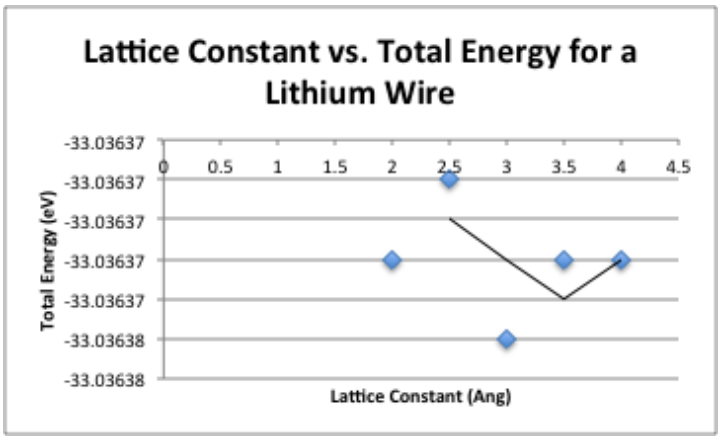


Figure 8*: Total energy for lithium wire apparatus based on changing angstrom values. The optimal value is 3.0 angstroms as seen by the minima, resulting in a structurally and eclectically realistic apparatus.*

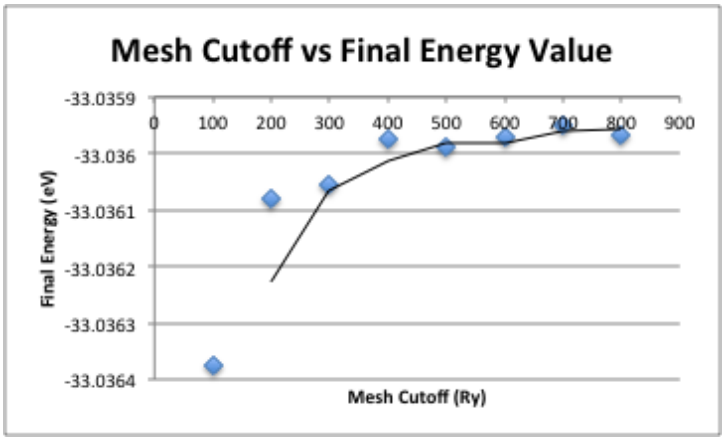


Figure 9: *Total energy for lithium wire apparatus based on changing mesh cutoffs. The optimal mesh cutoff is 700 rydbergs as seen by the convergence.*

Table VI. *Optimal conditions for a lithium wire. These conditions make sure the apparatus is spatially organized in the most logical and effective manner.*

	Lattice Constant	Mesh Cutoff
Optimal Conditions for Lithium Wire	3.0 Angstroms	700 Ry

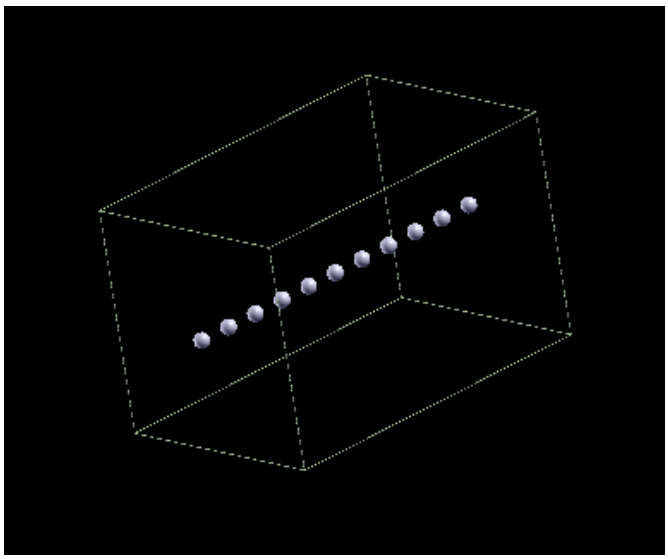


Figure 10: *Visual representation of lithium wire apparatus in a 16 by 16 by 40 angstrom unit cell.*

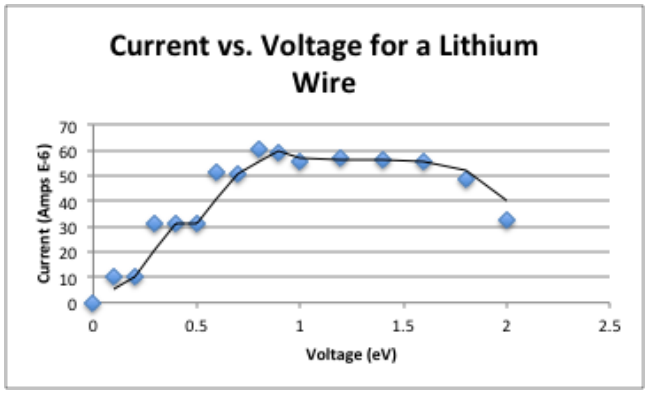


Figure 11: *Transistor like behavior displayed in the current vs. voltage behavior of the lithium wire apparatus*

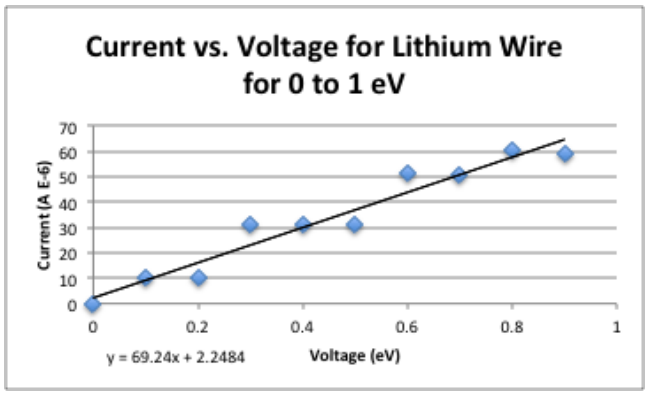


Figure 12: *Linear conductivity behavior of lithium wire apparatus when viewed from 0 to 1 electon volts.*

Aniline Structure and Conductivity-

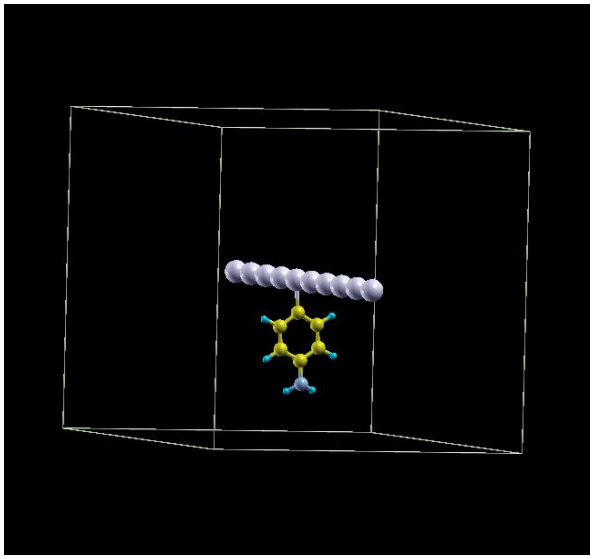


Figure 13: *Visual representation of aniline molecule attached to lithium wire apparatus in a 16 by 16 by 40 angstrom unit cell.*

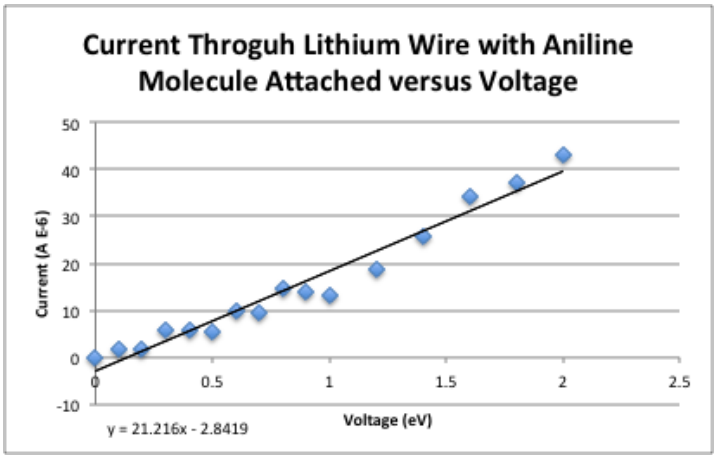


Figure 14: *Linear conductivity trend of current passing through aniline molecule attached to lithium wire apparatus.*

Benzene Structure and Conductivity-

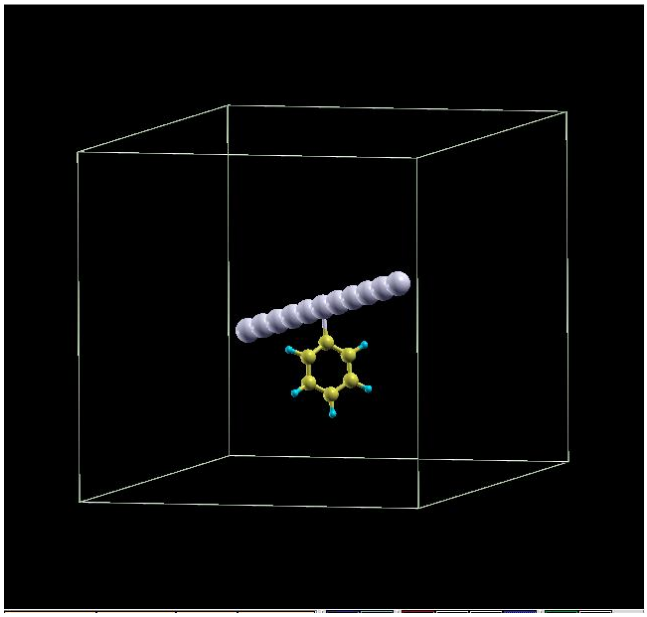


Figure 15: *Visual representation of benzene molecule attached to lithium wire apparatus in a 16 by 16 by 40 angstrom unit cell.*

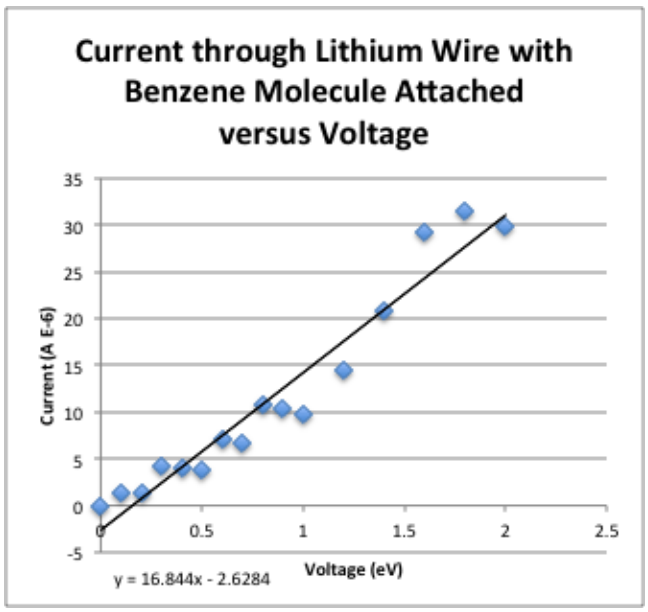


Figure 16: *Linear conductivity trend of current passing through benzene molecule attached to lithium wire apparatus*

Trichloroethylene Structure and Conductivity-

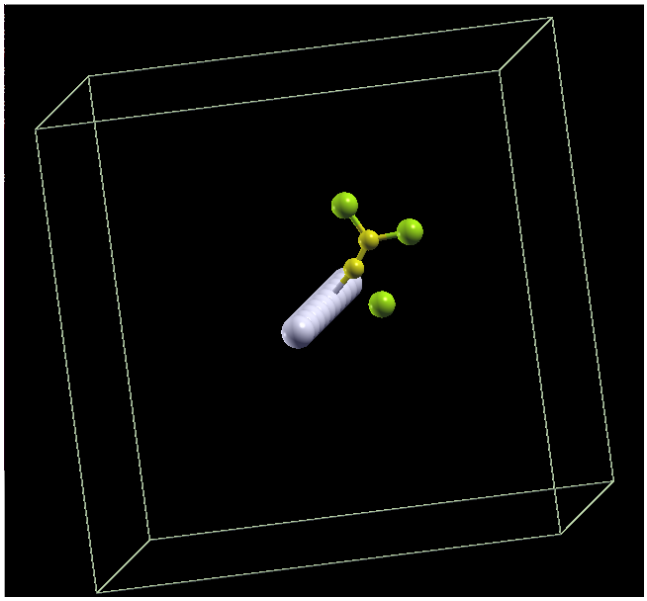


Figure 17: *Visual representation of trichloroethylene molecule attached to lithium wire apparatus in a 16 by 16 by 40 angstrom unit cell.*

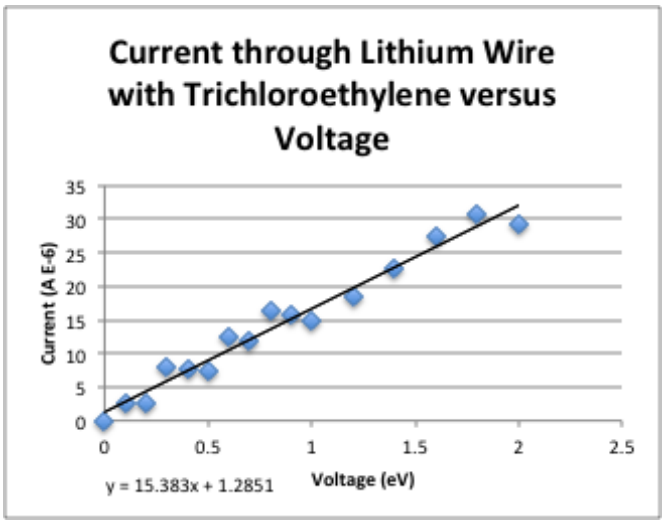


Figure 18: *Linear conductivity trend of current passing through trichloroethylene molecule attached to lithium wire apparatus.*

Table VII. Resistivity measurements of lithium wire apparatus with different test molecule attached. Attached molecules lead to increased resistance.

Apparatus	Resistivity (Ohms)
Lithium Wire	14.4k
Lithium Wire with Aniline Molecule	49.5k
Lithium Wire with Benzene Molecule	59.4k
Lithium Wire with Trichloroethylene Molecule	65.0k

Results and Discussion

Overall, the results indicate that a new type of either a resistor or a transistor device was created. Data from Figures 12, 14, 16, and 18 indicate that as voltage increases, current increases in a relatively linear fashion. This behavior is almost identical to transistor behavior, although no plateau is reached at a threshold voltage when a molecule is attached. However, the lithium wire does display this constant behavior as seen in Figure 11 at around 1 electron volt.

A trend seen in all the voltage testing and in Figures 12, 14, 16, and 18 is the step-like pattern of actual data points. While the overall trend is linear, multiple data points have the same or very similar current outputs. This can be related to the fact that energy is quantized on the quantum level. For example, the energy values for a particle in a one-dimensional box are:

$$E_n = \frac{n^2 h^2}{8ml^2}$$

with n being the eigenvalue which is restricted to positive integers only[26]. Therefore, the particle in a one-dimensional box can only take on discreet energy values very similar to how the lithium wire apparatus can only conduct certain amounts of current.

Table VI shows the optimal conditions for a lithium wire. The lattice constant for bonding lithium atoms is optimal around 3.0 angstroms. The minimum seen in total energy value in Figure 8 indicates that this distance produces the best bond. Additionally, the atomic radius of lithium is 1.47 angstroms[27], so 3.0 angstroms is a very logical distance for two bonding lithium atoms. Therefore, when two lithium atoms are 3.0 angstroms apart, there is a minimum in total energy due to structure, which is conducive to bond formation. When altering the mesh cutoff for the lithium wire, a large variety of final energy values are produced. However, as the mesh cutoff increases, the results start to converge to a smaller range around 700 rydbergs, as seen in Figure 9. This wide range of experimental final energy values follows the behavior of oscillating energies with increased mesh cutoff as observed in reality. Higher mesh cutoffs tend to converge to a particular final energy value. However, they also create a problem because large mesh cutoffs tend to cause higher computational times. With this in mind, 700 rydbergs offered an accurate final energy value without excessive calculation times[28]. Figure 10 shows that a lithium chain with sufficient spacing and 3D orientation was created using these parameters.

When voltage was applied to the optimized lithium chain, the conductivity patterns were very similar to transistor action[29]. Figure 12 shows that from 0 to 1 electron volts, the conduction pattern is very linear, aside from the stair-step pattern due to quantum effects. Figure 11 shows that from 1 electron volt to approximately 1.5 electron volts, the current conducted by the apparatus is constant. At higher voltages, however, the apparatus starts to

fail and lose current with increasing voltage. Since lithium is a metal, it should only conduct more electricity with higher voltages. Therefore the proceeding molecule attachment apparatuses were only tested up to 2 electron volts.

Before voltage testing began with molecules attached, Figures 13, 15, and 17 were generated in order to ensure that the proper structure was attained. Figure 17 shows that a bond did not form between a chlorine and carbon atom in the graphical program xcrysden. While the bond is not visualized in xcrysden due to similar z coordinates, it is important to note that the coordinates are close enough that SIESTA and tranSIESTA will treat the atoms as if they are bonded.

Aniline's band gap energy data follows an interesting pattern. It mimics both transistor like qualities with the linear growth on the left side of Figures 2 and 3, but instead of leveling off, the band gap energy drops rapidly. The data from Table III indicates that this drop in band gap energy occurs around 32 electron volts. This is a much higher voltage than what was tested when the aniline molecule was attached to the lithium wire. Therefore, it is hard to conclude whether the band gap energy of aniline had an effect on the current data received during the lithium chain testing. However, the linear, positive correlation between voltage and current appears in both the band gap data and voltage testing.

Data from Table II and Figures 4 and 5 indicate that the band gap energy of benzene greatly increases from 0 to 1 electron volts. Then, past 1 electron volt, the band gap energy becomes relatively constant. This behavior mirrors transistor action, but it is not reflected in the voltage testing with the lithium wire. Figure 16 shows a much more linear relationship between current and voltage with a spike in current occurring after 1 electron volt.

Band gap energy data from Table V and Figure 5 and 6 indicate another transistor like molecule. Although no threshold voltage is reached, the trichloroethylene band gap energy follows an extremely linear pattern. This is reflected in Figure 18. When attached to a lithium wire with voltage applied, current increases through the trichloroethylene apparatus in an extremely linear fashion. This highly resembles the band gap energy pattern, however all the other voltage data was linear as well. While it is possible that the band gap energy of trichloroethylene influenced the conductivity patterns of the trichloroethylene lithium wire apparatus, it is also likely that a lithium wire with any molecule attached has an intrinsically linear conductivity pattern.

It is also worth noting the data from Table VII. In all three cases, when test molecules were attached to the lithium wire, the resistivity of the apparatus increased significantly. This means less current was able to flow through the device when molecules were attached as opposed to a normal lithium wire. With this in mind, it also is possible that the data from Figures 14, 16, and 18 follow the ohmic behavior: the behavior of simple resistors[30]. Additional data must be gathered in order to determine whether the devices continue to display transistor like behaviors at higher voltages or simply act like resistors.

Overall, whether a unique resistor or transistor was created, the lithium wire apparatus is not feasible yet. Firstly, engineering has not reached the atomic level yet, so actually creating these devices would be extremely costly in terms of time and money. Additionally, implementing these devices in modern electronics would also be difficult. The electrodes on the apparatus consist of four atoms in a line so they would be very difficult to attach to anything.

Even with the engineering obstacles, molecular electronics similar to the ones designed in this paper are a worthwhile area of research. Right now, transistors are on average 45 nanometers long. That is 450 angstroms long, nearly 14 times longer than the lithium wire apparatus designed in this study[31]. Creating this new generation of molecular

transistors would have profound effects. Electronics could have 14 times the computational power while staying the same size.

Acknowledgements
I would like to thank Dr. Gary Washington for his lessons, guidance, and support throughout this research project.
I would also like to thank Mr. Gary Adelson for providing vital technical support to my research.

References
[1] Chandler, D. Explained: Band Gap http://news.mit.edu/2010/explained-bandgap- 0723 (accessed Jun 21, 2016).
[2] Yu, P.; Cardona, M.; Sham, L. Fundamentals Of Semiconductors: Physics And Materials Properties. *Phys. Today* 1997, *50*, 76.
[3] Derman, S. Modern Semiconductor Components And Circuits-The Basics. *The Physics Teacher* 1980, *18*, 619.
[4] Hoeneisen, B.Mead, C. Power Schottky Diode Design And Comparison With The Junction Diode. *Solid-State Electronics* 1971, *14*, 1225-1236.
[5] Costa, B.Cocito, M. Semiconductor Devices For Photonic Switching. *Eur. Trans. Telecomm.* 1991, *2*, 237-250.
[6] Brinkman, W.; Haggan, D.; Troutman, W. A History Of The Invention Of The Transistor And Where It Will Lead Us. *IEEE J. Solid-State Circuits* 1997, *32*, 1858-1865.
[7] Kane, E. Zener Tunneling In Semiconductors. *Journal of Physics and Chemistry of Solids* 1960, *12*, 181-188.
[8] Nave, C. Tunneling, Barrier Penetration http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/barr.html (accessed Jun 28, 2016).
[9] Haviland, D. The Transistor - History https://www.nobelprize.org/educational/physics/transistor/history/ (accessed Jul 5, 2016).
[10] Nave, C. Band Theory for Solids http://hyperphysics.phy-astr.gsu.edu/hbase/solids/band.html (accessed Jul 5, 2016).
[11] Rahman, Faiz. Solid-state Lighting with Wide Band Gap Semiconductors. *MRS Energy & Sustainability* 1 (2014). doi:10.1557/mre.2014.11.
[12]Boukhatem, M. H. *Carries Temperature Dependence of Energy Band Gap of Germanium: Silicon*2015, *8* (2), 309–312.
[13]Presanda, Ursman. "Intel ISSCC: 14nm All Figured Out, 10nm Is on Track, Moores Law Still Alive and Kicking." 2015. Accessed July 05, 2016. http://wccftech.com/intel-isscc-14nm/.
[14]Gabedit—A graphical user interface for computational chemistry softwares. Allouche, A.-R., Journal of Computational Chemistry, 32 (2011) 174–182. doi: 10.1002/jcc.21600
[15]M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, W.A. de Jong, MWChem: a comprehensive and scalable open-source solution for large scale molecular simulations" Comput. Phys. Commun. 181, 1477 (2010)
[16] Gilbert, A.; Hartree-Fock Theory*Computational Chemistry and Molecular Modeling* 93–113.
[17]Veillard, A. *Quantum Theory of Polymers*1978, 23–30.
[18]Shaik, S. *A Chemist's Guide to Valence Bond Theory* 81–93.
[19]José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal, *The SIESTA method for ab initio order-N materials simulation*, J. Phys. Condens. Matter **14**, 2745-2779 (2002).
[20]Kurt Stokbro, Jeremy Taylor, Mads Brandbyge, and Pablo Ordejón, *TranSIESTA: A Spice for Molecular Electronics*, Ann NY Acad Sci 1006, 212-226 (2003).

[21]Bergren, A. J.; Ivashenko, O. *Advances in Electrochemical Sciences and Engineering Alkire/Electrochemistry of Carbon Electrodes Electrochemistry of Carbon Electrodes*2016, 339–378.
[22]Bastawros, A. Crystal Structure http://www.public.iastate.edu/~bastaw/courses/mate271/week2.pdf (accessed Jul 2, 2016).
[23] Junquera, J. The eggbox effect: converging the mesh cutoff, 2016
[24]Washington, G. Coordinate Transfer, 2016.
[25]A. Kokalj, XCrySDen--a new program for displaying crystalline structures and electron densities, *J. Mol. Graphics Modelling*, 1999, *17*, 176--179.
[26] Bennett, D. Particle in a 1-dimensional box http://chemwiki.ucdavis.edu/core/physical_chemistry/quantum_mechanics/05.5:_particle_in_boxes/particle_in_a_1-dimensional_box (accessed Jul 8, 2016).
[27]Moore, J. *Hawley's Condensed Chemical Dictionary*2007.
[28]Hesselbarth, J.; Vahldieck, R. *1998 IEEE MTT-S International Microwave Symposium Digest (Cat. No.98CH36192)*.
[29]Cho, G. R.; Chen, T. *Fourth International Symposium on Quality Electronic Design, 2003. Proceedings.*
[30]Lampert, M. A.; Rose, A. *Phys. Rev. Physical Review*1959, *113* (5), 1236–1239.
[31]Pulfrey, D. L. *Understanding Modern Transistors and Diodes* 225–250.

How Effective is Fiscal Policy in Correcting Income Inequality?

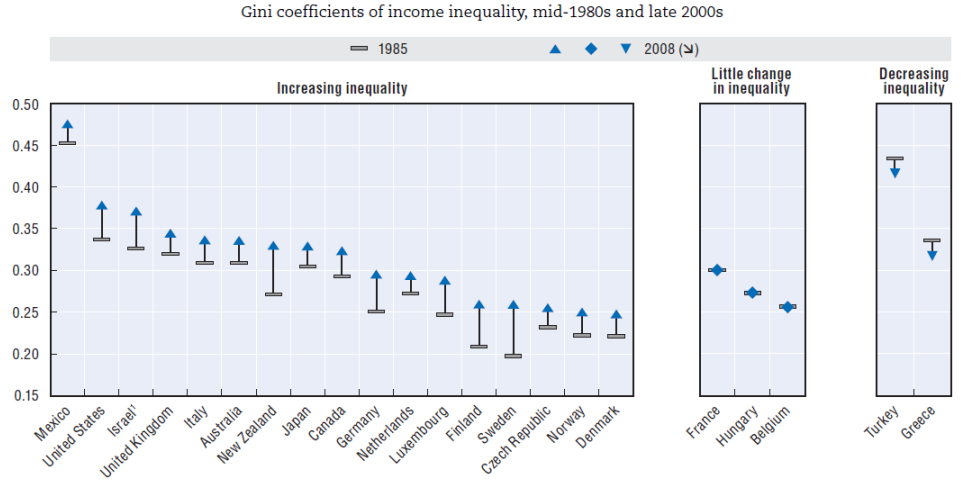
Tongxin Zhang

Author background: Tongxin Zhang grew up in China and currently attends WHBC of Wuhan Foreign Languages School, located in Wuhan, China. His Pioneer seminar topic was in the field of economics and titled "Monetary Policy and the Great Recession."

1. Introduction

Income equality has a close correlation with social well-being, including higher life expectancy, lower crime rates, and a stable economy (IMF, 2014). Thus, most governments have shown a growing attention to the distribution of income. Social spending, an expenditure targeted at low-income households, the elderly, and other vulnerable groups, stood at an average of 15.4 (percentage of GDP) in OECD (Organisation for Economic Co-operation and Development) countries in 1980. Thirty years later, it had increased to 21.7 (percentage of GDP), a percentage change of 41% (OECD, 2011).

However, there has been growing evidence that shows that income inequality has increased in most countries, as shown by the Gini coefficient, the standard measure of income inequality which has a value between 0 (perfect income equality) and 1 (perfect income inequality). The average Gini index for OECD countries had increased from 0.29 in the mid-1980s to 0.316 in 2008 which is a change of 10%(OECD, 2011). Most countries in OECD showed a large increase in inequality in the last three decades, as shown in Figure 1.



Note: For data years see Table 1. "Little change" in inequality refers to changes of less than 2 percentage points.
1. Information on data for Israel: http://dx.doi.org/10.1787/888932315602.
Source: OECD Database on Household Income Distribution and Poverty.

StatLink http://dx.doi.org/10.1787/888932535185

Figure 1

This increase in inequality can be attributed to several factors, including the " globalization and liberalization of factor and product markets; skill-biased technological change; increases in labor force participation by low-skilled workers; declining top marginal income tax rates; increasing bargaining power of high earners; and the growing share of high-income couples and single-parent households" (IMF, 2014, p. 5).

But there is good news: because of government's growing focus on income distribution, almost every country's Gini index shows a decrease after the implementation of relevant policies, as shown in Figure 2 (Mark Luebker, 2011). Among all those policies, fiscal policy becomes the primary and most widely-used by governments in correcting income inequality. Fiscal policy consists of two parts: government revenue including taxes, and government spending. Both tax and government expenditure can decrease the level of income inequality in the market. For example, income taxes can reduce the inequality of disposable income (income after taxes and transfers); similarly, unemployment benefits (a part of government spending) can contribute to income redistribution as well.

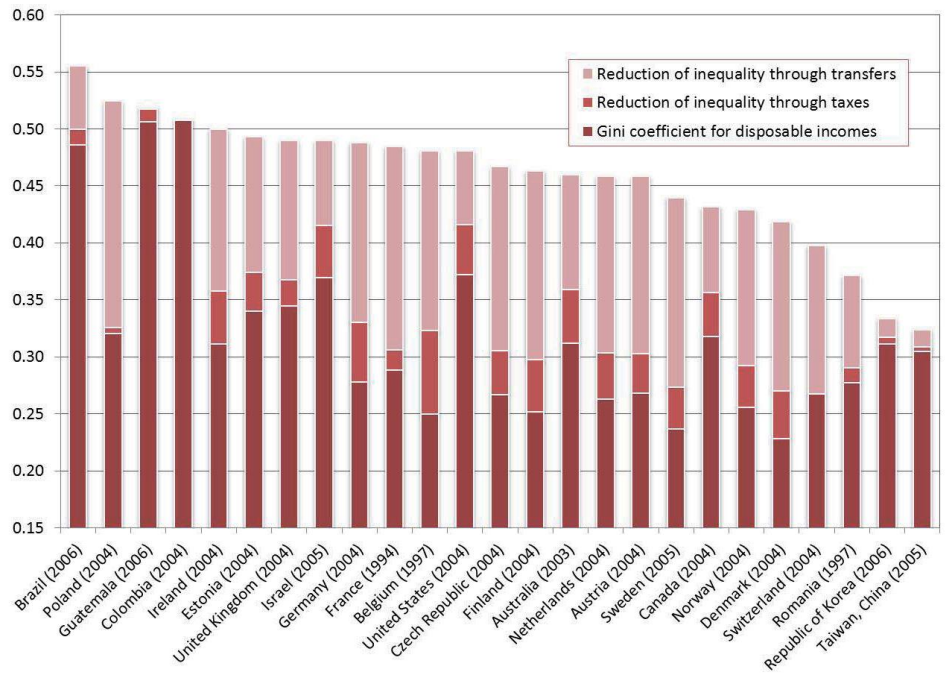


Figure 2

From Figure 2 above, it is clear that all the countries show a significant decrease in inequality after the use of fiscal policy. But the effects of fiscal policy on income inequality vary among those countries. Therefore, two questions come to mind: To what extent is fiscal policy actually effective in solving income inequality? Which factors determine the effectiveness of fiscal policy? The answers to those two questions can explain why the problem of inequality nowadays is still severe and how to really solve this problem.

2. Approaches to this question

In assessing the extent to which fiscal policy affects income distribution, I will consider two crucial parts of fiscal policy: government revenue (tax), and government expenditure. For government revenue, I will consider the progressivity and revenue of the taxes collected. For government expenditure, I will focus on the share of social spending as a percentage of the GDP.

In order to indicate the distribution of income in certain markets, I will use the Gini index and other measures including the Kakwani index, another measure of progressivity through social intervention. Kakawani index has a value range from -1 to 1 and the larger the index represents the larger progressivity. But the foremost most part of my paper will apply the Gini index before/after taxes and transfers (i.e. before and after fiscal policy) to show the effect of fiscal policy on income distribution.

2.1 Sample and data

For samples, I will investigate two countries in this paper:

- The United States;
- The United Kingdom.

I have chosen to analyze these two countries because the data for developed economies is more accessible and accurate than those of developing economies.

Between the two, the United Kingdom shows a large decrease in Gini index after tax/transfers. For the United States, the decrease in Gini index is relatively mild.

I will analyze these countries later and try to find an international comparison. The comparison between these two countries can help me find the key to my questions.

My data all come from 1985 to 2010, a period of 25 years, as I want to look for a long-term trend for each country. The timeframe cutoff is at 2010 because most of the data after that is missing and inaccurate, which would affect my investigation.

2.2 Limitations

The first limitation with this paper is that I am only analyzing developed countries. Though a more thorough analysis of the issue would include a comparison between developed and developing countries, I found that the scope of my paper was best limited to developed countries since there is insufficient data available for developing countries, such as Chile and Brazil.

Secondly, transfer payment, an important component of government spending, is not included in this paper because statistics about transfers are too hard to find online, so I choose social spending and cash benefits instead.

3. Analysis of fiscal policy in the United States

3.1 Overview - mild effect on inequality

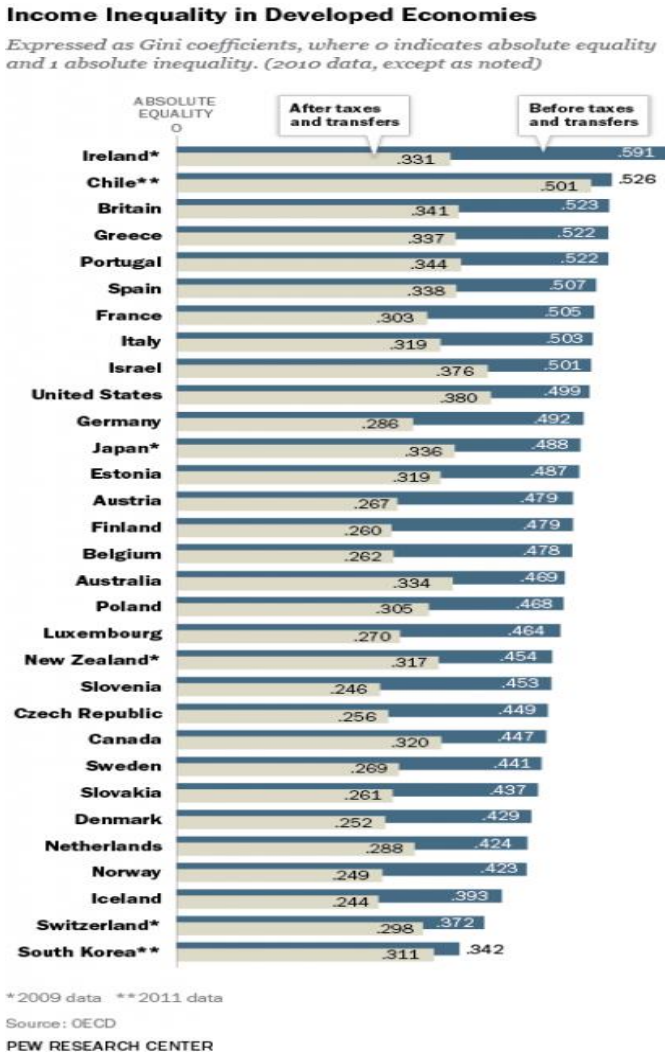


Figure 3

Before analyzing the United States directly, I will investigate how the United States compares internationally.

According to the data from OECD, which is a group of 34 countries, without taxes and transfers (that is, fiscal policy), the United States had a Gini coefficient of 0.480 and ranked 10th on the list of other developed countries that included European nations like France and Britain. However, after implementing fiscal policy, the United States ranked 2nd on the list, with a value of 0.380, only behind Chile, as shown in Figure 3 (Drew Desliver, 2013).Now it seems that the problem of inequality is not so severe in the United States, which only

ranked 10th on the list. What is severe is the problem with its fiscal policy's redistributive power, as the reduction in inequality was only 0.119 points (the reduction in Gini coefficient before/after fiscal policy), far behind Britain's 0.182 and France's 0.202.

Therefore, there must be something insufficient in the policy itself. Now let us find out. The next part of my paper seeks to find out what it is.

3.2 Tax system in the United States

3.2.1 How progressive is the United States' tax system?

Defined as a relation between income and the fraction of income paid as tax, taxes can be divided into three types: progressive, regressive, and proportional (Ellie Tragakes, 2014).

A tax system is called progressive when people with higher income pay higher taxes; here "higher" does not mean quantities, but percentage of income. In order to shrink the difference between the high and low income levels (and to reduce inequality), most countries have progressive tax systems.

As stated in "Economics for IB diploma" by Ellie Tragakes:
"[the] thing that can be said with certainty is that the more progressive a tax system, the more equal (or less unequal) the after-tax distribution of income becomes" (2014, p.313).

As illustrated in the previous section, the decrease in the United States' income inequality after fiscal policy was implemented is mild. So is this caused by a relatively regressive (or less progressive) tax system?

According to the research carried out by the OECD in 2008, the nation collected 45.1% of tax revenue from the richest 10%, far exceeding peer countries like Britain's 38.6%, Germany's 31.2%, or the OECD's average of 31.6%. Even considering the "ratio of shares on the richest 10%", the value 1.35 is much higher than Britain's 1.20, or the OECD's average 1.07. As shown in Table 1, the United States "has the most progressive tax system and collects the largest share of taxes from the richest 10% of the population" (OECD, 2008, p.106).

Measure of progressivity of taxes in selected OECD countries

Country	1. Share of taxes of richest decile (10%)	2. Share of market income of richest decile (10%)	3. Ratio of shares for richest decile (1/2)
France	**28.0**	**25.5**	**1.10**
Germany	**31.2**	**29.2**	**1.07**
Japan	**28.5**	**28.1**	**1.01**
United Kingdom	**38.6**	**32.3**	**1.20**
United States	**45.1**	**33.5**	**1.35**
OECD Average[21]	**31.6**	**28.4**	**1.11**

Table 1

21

http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

Even the tax collection on the richest 10% cannot prove the progressivity of the United States' tax system alone. Scott A. Hodge, the president of the Tax Foundation in Washington, wrote in 2008: "even after accounting for the fact that the top 10 percent of households in the U.S. have one of the highest shares of market income among OECD nations, our tax system is second only to Ireland in terms of its progressivity for households."

The results seem to be paradoxical: if the United States has such a progressive tax system, then why is the effect of fiscal policy still mild, and the after-tax income distribution still unequal?

The effect can be explained by the fact that Ellie Tragakes' argument is incomplete: the definition of progressive tax only specifies the increasing tax rate when income increases. It says nothing about the weight of taxes on people's income. Even if a country has 0.01% tax on its poorest 10% population and 0.8% on its richest 10% (which is a strongly progressive tax rate), the tax cannot have an obvious effect on inequality because the amount of tax collected is too small to play a role in income inequality.

The quantity matters.

3.2.2 How much tax revenue is collected in the United States?

In terms of the revenue collected, the result is not so satisfying. From 1980 to 2010, the United States' total tax revenue as percentage of GDP had always been less than the OECD average, and the gap between them grew continuously during that time period, from 1980's 4.6% to 2005's 8%, until 2010's 9.6%, as shown in Table 2.

Total Tax Revenue as percent of GDP over time						
Country	1985	1990	1995	2000	2005	2010
United States	24.6	25.9	26.4	28.2	25.9	23.2
OECD average[22]	31.5	32.1	33.6	34.2	33.9	32.8

Table 2

The United States' tax revenue is not only far below the OECD average, but lower than those of several peer countries. Take 2010 for example, the United States' 23.2% was much lower than Britain's 32.7% or France's 41.6%.

So the US's tax revenue is insufficient, too small to impose a substantial effect on income distribution. However, considering only total tax revenue (as percentage of GDP) is too generalized. It's necessary to analyze the two important parts of "total tax revenue": direct taxes and indirect taxes.

22

http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

3.2.3 Direct taxes

Personal income taxes, also known as "direct taxes," are the most important source of government tax revenues in many countries. Personal income taxes are progressive in the United States and can therefore reduce income inequality (Clive Crook, 2012).

However, with the exception of 2010, when the United States'10.5% exceeded the OECD's average of 9.2%, the United States' personal income tax revenue remained lower than the OECD average. And from 2000 to 2010, the weight of personal income tax on GDP continued to shrink, from 10.5% in 2000, to 8.3% in 2005, until it reached 7.1% in 2010 (Table 3).

Personal Income Tax Revenue as percent of GDP over time						
Country	1985	1990	1995	2000	2005	2010
United States	8.7	9.2	8.9	10.5	8.3	7.1
OECD - Average[23]	9.9	10.3	9.2	9.2	8.7	8.3

Table 3

23 http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

Year	Average income taxes rate
1985	7.8
1986	7.7
1987	8.1
1988	7.7
1989	8.1
1990	8.0
1991	7.6
1992	7.4
1993	7.5
1994	7.6
1995	7.8
1996	8.3
1997	8.7
1998	9.1
1999	9.2
2000	9.6
2001	9.2
2002	7.5
2003	6.7
2004	6.5
2005	7.1
2006	7.5
2007	8.0
2008	7.9
2009	6.0
2010	6.2

Table 4

Even if the average personal income tax rate (as percent of gross income) were considered, the US shows a decreasing trend over time: from 1985's 7.8% to 2010's 6.2%. Even as the value reached 8.0% in 2007, it kept decreasing in the next three years (Table 4).

Therefore, both income tax revenue and the income tax rate show that the United States is doing an insufficient job in collecting direct taxes.

3.2.4 Indirect taxes

Sales taxes are taxes on goods or services, also known as "indirect taxes" (Ellie Tragakes, 2014). In the United States, sales taxes are fixed percentages of the retail prices of goods and services. No matter who buys the goods or services, the taxes paid are the same. Therefore, sales taxes are regressive, as they are a fraction of income paid, and as income increases, the percentage of it used to pay the tax decreases; sales taxes have a negative effect on income distribution. Nearly every country's Gini coefficient will increase after the implementation of indirect taxes.

If I compare the sales taxes (as a percent of GDP) in the United States to the OECD-average, the results are much more promising. From 1985 to 2010, the United States' taxes on goods

and services were less than half of the OECD-average. This time, less is better, as sale taxes are regressive. (Table 5)

The United States is collecting a much lower amount of indirect taxes, therefore with less increase in inequality level (after indirect taxes).

Taxes on goods and services as percent of GDP						
Country	1985	1990	1995	2000	2005	2010
USA	4.631	4.573	4.796	4.554	4.594	4.306
OECD-average[24]	10.161	10.165	11.106	11.013	10.994	10.696

Table 5

3.2.5 Summary on the United States' tax system

The United States' tax system is very progressive, as the richest 10% pay the highest share of taxes compared to other countries in the world: 45.3%. However, as Chye-Ching Huang and Nathaniel Frentz (2014) point out in their report "What Do OECD Data Really Show About U.S. Taxes and Reducing Inequality?": "the amount of revenue a tax system raises, as well as its progressivity, determines how effectively the tax system reduces income inequality."When it comes to tax revenue, the nation falls behind the OECD average.

From the analysis of personal income taxes and sales taxes, the United States underuses progressive taxes, with a decreasing tax rate over time. But the good news is that the United States collects much less regressive indirect taxes, less than half of the OECD average value. The two types of taxes seem to cancel each other out.
However, problems still exist: primarily, the low tax revenue collected. Low tax revenue leads to multiple effects: as tax revenue remains low, the government budget is affected, and government cannot spend much on transfer payments or social expenditure to further solve inequality.

Then comes the government spending.

3.3 Government expenditure in the United States

3.3.1 Overview

After evaluating the tax system in the United States, I found that there is something insufficient about this system: either the amount of tax collected is too low, or the regressive sales taxes are overused. These problems in the tax system have impacted the effectiveness of its fiscal policy to some extent.

24 http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

However, tax collection is not the only part of fiscal policy, as fiscal policy also includes government expenditure. In this chapter, I will find out whether the US government has spent enough money on solving income inequality.

3.3.2 Evaluation of social spending

Social spending includes cash benefits, provision of goods/services, and tax breaks for the underclass. The targets of social expenditure may be low-income households, elders, disabled or unemployed people in the society. Described as "social", these kinds of spending must have a redistributive effect on household resources including income (OECD, 2014).

Among numerous branches of government spending, social spending is the one that is targeted directly at diminishing inequality, so I use this index (as percent of GDP) for evaluation.

USA	Gini (disposable income, post taxes and transfers)	Gini index (Market income, before taxes and transfers)	Difference in income inequality	Social spending as percent of GDP
1985	0.340	0.436	0.096	12.80%
1990	0.349	0.450	0.101	13.10%
1995	0.361	0.477	0.116	14.00%
2000	0.357	0.476	0.119	14.20%
2005	0.380	0.486	0.106	15.50%
2010	0.380	0.499	0.119	19.30%

Table 6

According to the data on the Gini index before/after fiscal policies, whenever social spending (as percentage of GDP) increased, the difference in inequality increased with the exception of the time period between 2000 and 2005. As a difference in inequality can represent the effect of fiscal policy, there exists a positive relationship between social spending and the effect of fiscal policy (as it can be seen from Table 6).

What is more optimistic is that the US government puts more emphasis on social expenditure, increasing from 12.80% in 1985 to 19.30% in 2010, a change of 51%. The nation's way of solving inequality is heading in a good direction.

However, when compared to the OECD-average, the United States' social spending seems insufficient. From 1985 to 2010, the United States' social spending had always been behind the OECD average value, but the gap is at least shrinking: from 2000's 4.41% to 2010's 2.36% (Table 7).

Social Spending as percent of GDP						
Country	1985	1990	1995	2000	2005	2010
USA	12.80%	13.10%	14.00%	14.20%	15.50%	19.30%
OECD-average[25]	17.01%	17.47%	19.30%	18.61%	19.42%	21.66%

Table 7

3.3.3 Evaluation of cash benefits

Cash benefits are a major component of social spending. Cash benefits mean payments to vulnerable groups including the unemployed and elderly. Like social spending, cash benefits always serve the purpose of decreasing inequality (while decreasing poverty).

According to the OECD's 2008 report "Growing Unequal," in the mid-2000s, US cash benefits were 9.4% of household disposable income, ranked the second lowest in the OECD, only better than Mexico's 5.8%, and far behind the OECD average of 21.9%.

The progressivity of the United States' cash benefits was -0.089 (measured in concentration coefficient, the lower, the more progressive), still lower than OECD's average value -0.099 (OECD, 2008).

The same report shows that because of their relatively small volume, the US cash benefits did not have a substantial effect on income inequality, with only 0.041 inequality reduction, much less than the OECD's average 0.078 reduction. Among 24 OECD countries, the US's effectiveness of cash benefits is the second lowest, only better than South Korea (OECD, 2008).

3.3.4 Summary on the United States' government expenditure

In terms of social spending, the United States shows a promising trend: an increase in social spending (as percent of GDP) of 51% in 25 years, and the gap between the United States' social spending and OECD average is shrinking. If the United States can continue placing the emphasis on social spending, it will quickly reach the OECD standard.

For cash benefits, the relatively small size and low progressivity limit their effect on solving income inequality. As the OECD's 2008 report points out, cash benefits have shown effective progress in reducing inequality for other countries. Therefore the United States should put more emphasis on cash benefits in the future (OECD, 2008).

In this section, I described the effect of the US's fiscal policy as "mild." However, if I consider the international comparison in 2010, the United States' effectiveness is only better than a few OECD countries.

The next section will introduce the country whose fiscal policy does a much better job than the United States: the United Kingdom.

---

25 http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

4. The United Kingdom

4.1 Overview

Just as I did with the USA, I want to first compare the United Kingdom internationally. According to the 2010 data (Figure 3), the United Kingdom had a Gini coefficient of 0.523 before fiscal policy, ranked 3rd among OECD countries. After fiscal policy, the United Kingdom's level of inequality decreased to 0.341, ranked 5th, better than the United States and Portugal (Drew Desilver, 2013).

It may not be effective to show the impact of the UK fiscal policy through improvements in ranks, since it only moved up two places. Instead, I consider the decrease in inequality after fiscal policy. The Gini coefficient decreased to the value of 0.182 after fiscal policy, much higher than the OECD average 0.158 decrease. Although in Europe, the effects of the UK's fiscal policy were still weaker than in France or Germany (both with 0.20^{+} decrease), its effectiveness was better than in most OECD countries.

If I compare the UK's fiscal policy to that of the United States from 1985 to 2010 (Table 5), some interesting patterns are revealed. Before taxes/transfers, the inequality level in the UK was much more severe, with over 0.5's Gini coefficient, than in the USA; however, after fiscal policy, its inequality level became much better than in the USA. The average decrease in inequality for the UK is much higher than the USA (0.163 VS 0.110), so the UK's fiscal policy imposes a much more substantial effect on income inequality than the USA. So why is this happening? (Table 8)

UK	Gini (disposable income, post taxes and transfers)	Gini index (Market income, before taxes and transfers)	Difference in income inequality
1985	0.309	0.469	0.160
1990	0.355	0.490	0.135
1995	0.337	0.507	0.170
2000	0.352	0.512	0.160
2005	0.335	0.503	0.168
2010	0.341	0.523	0.182

Table 8

4.2 Tax System in the UK

4.2.1 How Progressive is the tax system in the UK?

As shown in Table 1, the 2008 data, the share of taxes for the richest 10% in the UK is about 38.6% of the overall tax revenue, with their share of market income at 32.3%. For the ratio of shares for the richest 10%, the UK has a value of 1.20, better than the OECD average 1.11 (OECD, 2008).

The ratio 1.11 ranked 6th, and the value 38.6% ranked 4th among 24 OECD countries. Although not as progressive as the United States, the UK's tax system is fair, placing the country in the top half of the rankings (OECD, 2008).

4.2.2 How much tax revenue is collected in the UK?

After considering the progressivity of the UK's tax system, I then analyzed the exact amount of taxes collected in the UK.

From 1980 to 2010, the UK's overall tax revenue as a percent of GDP has always been higher or closer to the OECD average value (as in 1995 or 2010). If compared to the United States, the UK's taxes collected are much higher than the US's, with the gap as large as 9.6% in 2010 (seen from Table 9).

Tax revenue as percent of GDP							
Country	1980	1985	1990	1995	2000	2005	2010
United Kingdom	33.4	35.1	32.9	31.9	34.7	33.8	32.8
United States	25.5	24.6	25.9	26.4	28.2	25.9	23.2
OECD - Average[26]	30.1	31.5	32.1	33.6	34.2	33.9	32.8

Table 9

Because of the amount of revenue collected, the United Kingdom's taxes can impose a much more substantial effect on inequality through fiscal policy. However, only considering the UK's tax system in general terms is too limited. As in the previous analysis of the United States, I will separate the UK's taxes into two parts: "direct taxes" and "indirect taxes".

26 http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

4.2.3 Direct taxes

UK 's Effect of direct taxes on income distribution, 1985-2010			
Year	Progressivity (measured in Kakwani index)	Average rate (as percent of gross income)	Redistributive impact (percent reduction in Gini)
1985	12.9	21.6	-9.5
1986	11.9	21.3	-8.4
1987	11.6	21.4	-8
1988	8.9	20.8	-5.4
1989	8.7	20.2	-5.2
1990	8.4	20.3	-4.7
1991	10.2	19.5	-5.8
1992	12.2	19.4	-7
1993	13.3	19.5	-7.9
1994	14.2	20.5	-8.6
1995	15.9	20.6	-9.6
1996	14.1	19.8	-8.5
1997	14.3	19.9	-8.7
1998	13.4	20.7	-8.3
1999	12.6	20.4	-7.5
2000	12.2	20.5	-7.7
2001	12.8	20.6	-7.9
2002	15.8	19.7	-9.6
2003	15	20.7	-9.6
2004	15.2	20.8	-9.9
2005	15.1	20.9	-9.6
2006	13.5	21.4	-8.7
2007	13.8	21.4	-9
2008	14.2	20.5	-8.8
2009	15.6	20.2	-10
2010	14.7	20.1	-8.7

Table 10

According to the data from the Office for National Statistics (Table 10), the UK's taxes have been progressive at all times between 1985 and 2010 based on the measures of Kakwani index. Although there have been several fluctuations during this time (like 1995-1996, or 2009-2010), the overall trend showed an increase in progressivity (Office for National Statistics, 2016).

Average tax rates (as percent of gross income) is another story. Although the changes in tax rate had been steady, with few fluctuations during the 30 years, the overall trends showed a decrease in tax rates, from 1985's 21.6 to 2010's 20.1.

Personal Income Tax Revenue as percent of GDP over time						
Country	1985	1990	1995	2000	2005	2010
United Kingdom	9.1	9.7	9.2	10.2	9.8	9.4
OECD - Average[27]	9.9	10.3	9.2	9.2	8.7	8.3

Table 11

The same trend can be found if I consider income taxes as a percentage of GDP: income tax revenue decreased from 2000's 10.2% to 2010's 9.4%, although still above the OECD average. The overall trend showed a decrease in tax revenue (Table 1).

On the one hand, then, direct taxes are becoming more progressive and more targeted; on the other hand, the average rate is decreasing. Progressivity and revenue amounts are two factors determining the effectiveness of taxes in solving inequality, using the UK's direct taxes. However these two factors are moving in opposite directions. This can explain why the redistributive effect of direct taxes has been steady over time, with a value around -9% or -10%, as shown in Figure 4 (Office for National Statistics, 2016).

Progressivity, average rate and redistributive impact of direct taxes, 1977 to 2012

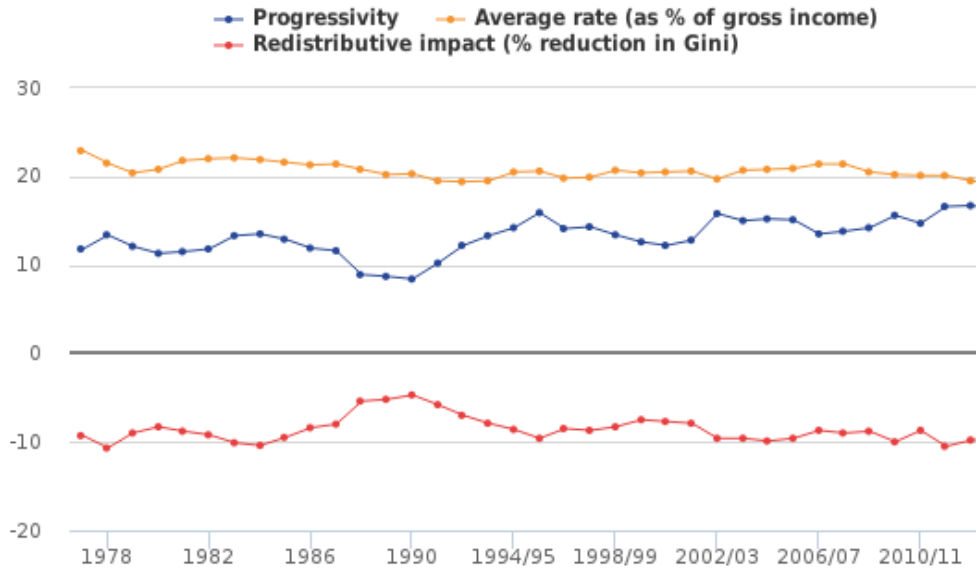


Figure 4

27

http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

4.2.4 Indirect taxes

The UK 's Effect of indirect taxes on income distribution, 1985-2010			
Year	Progressivity (measured in Kakwani index)	Average rate (as percent of gross income)	Redistributive impact (percent increase in Gini)
1985	-5.6	21.3	8.5
1986	-7.5	21.7	10.0
1987	-9.1	20.3	9.7
1988	-10.3	20.5	9.7
1989	-9.7	20.1	9.3
1990	-11.7	19.1	9.3
1991	-11.3	19.8	9.9
1992	-11.4	20.0	10.1
1993	-11.5	20.2	10.4
1994	-11.3	19.9	10.1
1995	-11.3	20.7	11.1
1996	-12.1	20.1	10.5
1997	-11.6	20.2	10.3
1998	-11.5	20.6	10.2
1999	-13.5	20.2	10.5
2000	-12.6	20.7	10.9
2001	-15.8	18.9	11.0
2002	-13.0	19.1	10.2
2003	-13.7	19.0	10.9
2004	-12.8	18.7	10.6
2005	-13.0	18.3	10.0
2006	-14.9	18.1	10.9
2007	-14.6	17.7	10.2
2008	-13.9	16.7	9.5
2009	-15.0	16.2	10.1
2010	-15.3	17.3	10.9

Table 12

As the data from the Office for National Statistics also shows (Table 12), the UK's indirect taxes have been regressive for all 25 years, shown in the negative data of the Kakwani index. The overall trends for indirect taxes showed a steadily decreasing value, from 1985's -5.6 to 2010's -15.3, a large drop of 273%. This is not good news as it shows that the UK's indirect taxes are becoming more and more regressive, and less targeted.

However, the UK’s indirect tax rate showed an opposite pattern. From 1985 to 2010, indirect tax rate decreased from 21.3% to 17.3%. These changes in tax rate are satisfying,

showing that the UK government is using less regressive taxes over time (Office for National Statistics, 2016).

Taxes on goods and services as percent of GDP						
Country	1985	1990	1995	2000	2005	2010
United Kingdom	11.00	10.200	11.240	11.050	10.220	10.110
OECD-average[28]	10.161	10.165	11.106	11.013	10.994	10.696

Table 13

The result is also optimistic if taxes on goods and services (a large composition of indirect taxes) are considered: such taxes used to be above the OECD average (from 1985 to 2000). However, they showed a decrease from 11.050 to 10.110 from 2000 to 2010, now below the OECD average value (Table 13).

Considering the UK's indirect taxes in general, the same thing happens again: progressivity, and taxes rate move in opposite direction. The redistributive effect again shows a steady value of around 10% or 11% (the value is positive as indirect taxes worsen income inequality), as shown in Figure 5.

Progressivity, average rate and redistributive impact of indirect taxes, 1977 to 2012

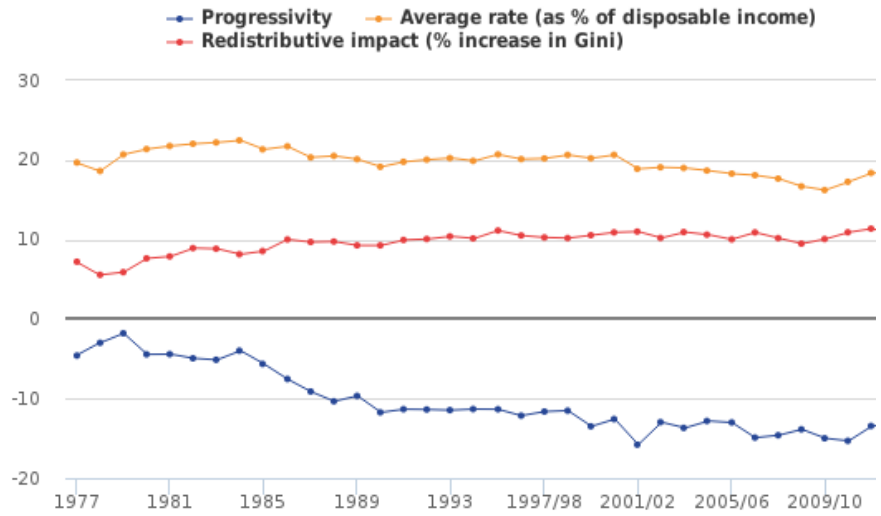


Figure 5

28

http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

4.2.5 Summary on the UK's tax system

In terms of total tax revenue raised, the UK does a very good job, staying above the OECD average, which is far better than the United States.

However, after analyzing the two crucial components of the UK's tax system, it is clear that both direct and indirect taxes have steady effects on income inequality over time, with the progressivity and revenue amounts moving in opposite directions for both types of taxes. Direct taxes decrease inequality about 9% or 10%, while indirect taxes increase inequality about 10% to 11%.

Changes in Gini index from taxes and spending, 1977 to 2012

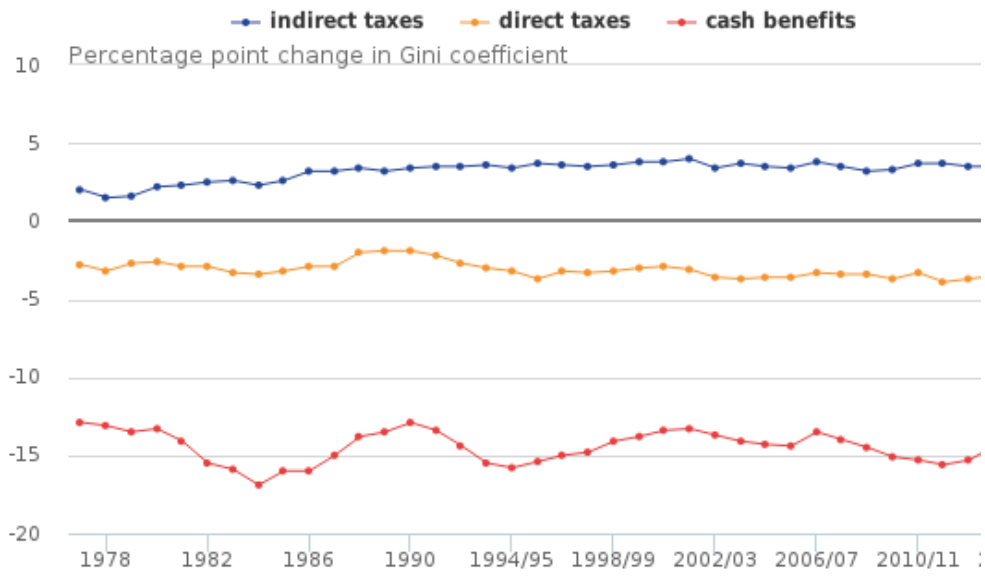


Figure 6

The two taxes' effects on income distribution are so close in absolute values that they nearly cancel out in total, showing no obvious effect on income inequality. What really shows a large effect on income inequality are cash benefits rather than taxes, as shown in Figure 6 (Office for National Statistics, 2016).

Cash benefits outlined in Figure 6 represent a kind of government expenditure - another part of fiscal policy. If taxes cannot work, then government spending may become a solution.

4.3 Government Spending in the UK

4.3.1 Overview

After completing the chapter on the UK's tax system, I discovered that the effect of taxes on income distribution is fairly small, as indirect taxes and direct taxes cancel out each other. However, as Figure 6 shows, cash benefits decrease Gini index over 15%.

For the UK, therefore, government spending is the key.

4.3.2 Evaluation on social spending

According to the OECD data about social spending (as a percentage of GDP), the UK's social spending suffered from an increasing trend from 1985 to 2010, from 19.20% to 22.80%; the effectiveness of fiscal policy also suffered from a large improvement, from 0.160 in 1985 to 0.182 in 2010 (Table 14).

UK	Gini (disposable income, post taxes /transfers)	Gini index (Market income, before taxes/ transfers)	Difference in income inequality	Social spending as percent of GDP
1985	0.309	0.469	0.160	19.20%
1990	0.355	0.490	0.135	16.30%
1995	0.337	0.507	0.170	19.20%
2000	0.352	0.512	0.160	18.40%
2005	0.335	0.503	0.168	20.20%
2010	0.341	0.523	0.182	22.80%

Table 14

When compared to the average value for OECD, the UK's social spending is higher than OECD-average or very close to the average value (except in 1990). The data about social spending indicates that the UK's extent of government expenditure is in the top half of the rankings among 24 OECD countries (Table 15).

Social Spending as percent of GDP						
Country	1985	1990	1995	2000	2005	2010
UK	19.20%	16.30%	19.20%	18.40%	20.20%	22.80%
OECD-average[29]	17.01%	17.47%	19.30%	18.61%	19.42%	21.66%

Table 15

29 http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REV&Coords=%5bCOU%5d.%5bOAVG%5d&ShowOnWeb=true&Lang=en.

4.3.3 Evaluation on Cash Benefits

UK 's Effect of cash benefits on income distribution, 1985-2010[30]			
Year	Progressivity(measured in Kakwani index)	Average rate (as percent of original income)	Redistributive impact (percent reduction in Gini)
1985	-91.8	19.4	-32.1
1986	-93	18.9	-31.7
1987	-94	17.5	-29.4
1988	-94	15.5	-27.1
1989	-94.4	15.1	-27
1990	-95.7	14.2	-24.9
1991	-94.1	15.2	-26.2
1992	-93.5	17.4	-27.9
1993	-95.3	18.6	-29.1
1994	-96.3	18.4	-29.9
1995	-94.6	18.4	-29.8
1996	-97	17.3	-28.5
1997	-97	16.8	-28.1
1998	-98.8	15.8	-26.8
1999	-97.5	15.5	-26.3
2000	-97.4	15	-26.1
2001	-98.3	14.8	-25.3
2002	-96.5	15.7	-26.8
2003	-96.3	16.2	-27.3
2004	-96.4	16.2	-28.2
2005	-96	16.6	-27.6
2006	-96	15.7	-26.2
2007	-96	16.2	-27.1
2008	-95.4	17.2	-27.8
2009	-94.5	18.5	-29
2010	-93.7	18.6	-29.3

Table 16

For cash benefits, with progressivity measured by the Kakwani index, a negative value means that cash benefits are progressive. Although they had an increase in progressivity from 1985 to 1998, cash benefits have become less progressive and less targeted since 1999 (Office for National Statistics, 2016).

In terms of the average rate, the cash benefits have constituted a larger and larger portion of people's income since 2000s, with the average rate increasing from 14.8% to 18.6%. This

led to an increasing trend for redistributive impact of cash transfers, from 2001's 25.3% to 2010's 29.3% (Table 16).

If I compare cash benefits to taxes, the average rate is even less than that of direct taxes and indirect taxes. However, the loss in amount of revenue is corrected through large progressivity (around -100). Therefore, cash benefits can have a large impact on the UK's fiscal policy, reducing the Gini coefficient by around 30% (Figure 7).

Changes in Gini index from cash benefits, 1977 to 2012

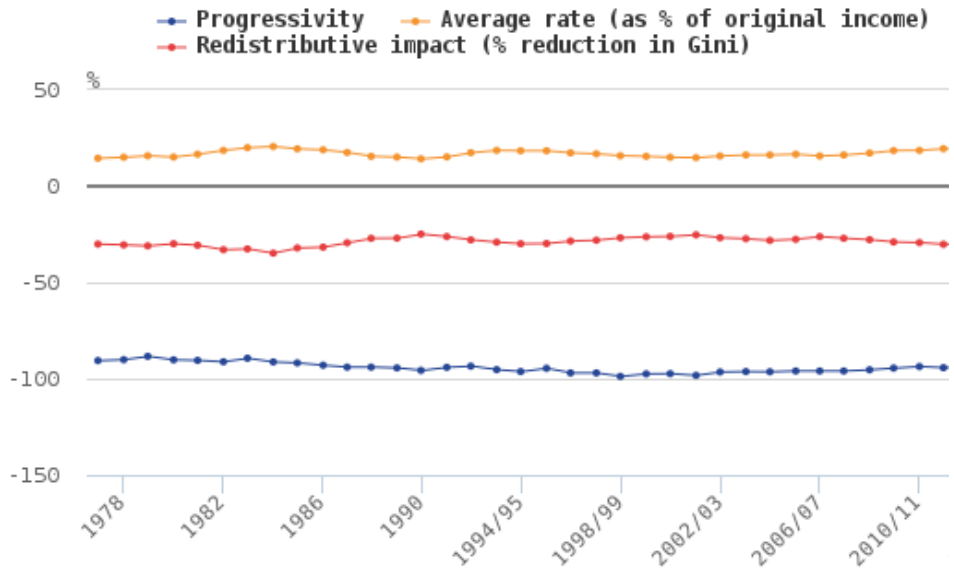


Figure 7

As shown in the 2008 OECD report "Growing Unequal?", the UK's cash benefits have a progressivity of -0.275 measured in concentration coefficient, far better than the OECD average value of -0.099, making them the most progressive cash benefits only behind Australia.

In the same report, the UK's cash benefits have an average rate of 22.7%, above OECD average of 21.9%. As the UK's cash benefits are both better in progressivity and amount, the country's effectiveness in reducing inequality using cash benefits is measured as 0.087, still above the OECD average value of 0.078(reduction in concentration coefficient) (OECD, 2008).

4.3.4 Summary on the UK's government spending

The UK shows a much higher emphasis on social spending, above the OECD average most of time from 1985 to 2010.

As to cash benefits, although the average rate is only a little behind the OECD average value, the strong progressivity in cash benefits makes cash benefits the strongest part of the UK's fiscal policy, reducing inequality by over 30%.

5. Future recommendations for the two countries

For the United States, as shown in Table 5, its fiscal policies have a mild effect on inequality. In order to improve the effectiveness of fiscal policies, several things should be noted:

- A progressive tax system is not enough. The amount of the revenue raised matters as well;
- Make more use of progressive taxes like personal income tax. Keep reducing the part played by regressive taxes like the sales tax;
- The effects of cash benefits on income inequality are substantial, and the government should utilize this tool more often.

For the UK, in order to have a more effective fiscal policy to reduce income inequality, several things need to be noted:

- Keep the progressivity of direct taxes, but increase the average rate collected;
- Make indirect taxes more progressive;
- Maintain the high progressivity of cash benefits, and increase their size a little more.

6. Conclusion

After analyzing the cases of the United States and the United Kingdom, I found out that both taxes and government spending play important roles in reducing income inequality in implementation of fiscal policy.

Although the United States has a high progressivity of taxes, the nation collects too little revenue. Therefore taxes have no substantial impact on income distribution. Although scoring less in progressivity, the UK collects enough revenue from taxes, and therefore has a more effective fiscal policy on income distribution (Table 17).

	USA	UK
Progressivity of Taxes	0.59	0.53
Amount of Taxes as percent of GDP	23.20%	32.80%
Progressivity of Cash Benefits	-0.089	-0.275
amount of Cash Benefits as percent of gross income	9.40%	22.70%
Social Spending as percent of GDP	19.30%	22.80%

Table 17

But taxes are not everything. US government spending is lower and less targeted than that of the UK (this analysis includes both social spending and cash benefits). Insufficient uses of both taxes and expenditure lead to the USA's dilemma.

How effective is fiscal policy in reducing income inequality? First, a progressive tax system is the basis for an effective fiscal policy. Second, in order for taxes to have a deep impact on income distribution, the amount of taxes collected matters, too. Third, progressivity and the amount of government spending matter, including social spending and cash benefits. As seen in the case of the UK, government spending is the key to its success in reducing income inequality. Lastly, both countries could substantially improve their income inequality with a combination of more progressive direct taxes and fewer regressive indirect taxes.

7. Bibliography

Anderton, A. (2008). *Economics (Fifth Edition).* Causeway Press.

Brian, N. (2009). Comprehensive Dictionary of Economics. ABHISHEK PUBLICATIONS.

DeSilver, D. (2013, December 19). *Global inequality: How the U.S. compares*. Retrieved from Pew Research Center website http://www.pewresearch.org/fact-tank/2013/12/19/global -inequality-how-the-u-s-compares

Growing unequal?: Income distribution and poverty in OECD countries. (2008). Paris: OECD, Organisation for Economic Co-operation and Development.

Hodge, S. A. (2008, October 29). *News To Obama: The OECD Says the United States Has the Most Progressive Tax System.* Retrieved from Tax Foundation website http://taxfoundation.org/blog/hnews-obama-oecd-says-united-states-has-most-progressive-tax-system

International Monetary Fund (IMF), (2014, January 23). *Fiscal Policy and Income Inequality*, Retrieved from http://www.cc.gatech.edu/user_surveys/survey-1998-10/

Luebker, M. (2011, August 05). *The impact of taxes and transfers on inequality*, Retrieved from International Labor Organization (Ilo) website http://www.ilo.org/travail/whatwedo/publications/WCMS_160436/lang--en/index.htm

Organization for Economic Co-operation and Development (OECD), (2011). *An Overview of Growing Income Inequalities in OECD Countries: Main Findings*,. Retrieved from https://www.oec d.org/els/soc/49499779.pdf
Parkin, M. (2010). *Macroeconomics.* Englewood Cliffs, NJ: Prentice-Hall.

Personal current taxes: Federal: Income taxes. (2015, August 10). Retrieved from Federal Reserve Bank of St. Louis (FRED) website https://fred.stlouisfed.org/series/B231RC1A027 NBEA#0

Revenue Statistics-OECD countries: Comparative tables. (2016). Retrieved from Organization for Economic Co-operation and Development (OECD) website https://stats.oecd.org/Index.as px?DataSetCode=REV

Social protection-Social spending-OECD Data. (2014). Retrieved from Organization for Economic Co-operation and Development (OECD) website https://data.oecd.org/socialexp/social-s pending.htm#indicator-chart

Statistical bulletin: The effects of taxes and benefits on income inequality: 1977 to financial year ending 2015. (2016). Retrieved from Office for National Statistics website https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/theeffectsoftaxesandbenefitsonincomeinequality/1977tofinancialyearending2015

Tax-Tax on goods and services-OECD Data. (2015). Retrieved from Organization for Economic Co-operation and Development (OECD) website https://data.oecd.org/tax/tax-on-goods-a nd-services.htm#indicator-chart

Tax-Tax on personal income-OECD Data. (2015). Retrieved from Organization for Economic Co-operation and Development (OECD) website https://data.oecd.org/tax/tax-on-personal-incom e.htm#indicator-chart
Tax-Tax revenue-OECD Data. (2015). Retrieved from Organization for Economic Co-operation and Development (OECD) website https://data.oecd.org/tax/tax-revenue.htm#indicator-chart

Tragakes, E. (2009). *Economics for the IB Diploma.* Cambridge: Cambridge University Press.

Study of Neural Circuits Involved in the Intuitive Decision Making Process in Teleostei

Pranav Bharat Khemka

Author background: Pranav Bharat Khemka grew up in India and currently attends Jamnabai Narsee International School, located in Mumbai, India. His Pioneer seminar topic was in the field of neuroscience and titled "The Decision Making Brain."

Abstract

Reflex actions play a vital role in an animal's survival. Oftentimes, stimuli do not reach the cerebrum or the cerebellum and decisions are made by intermediate neuronal circuits. The Mauthner cell system is responsible for controlling the C-start (see Glossary) escape in teleostei, and it has been widely studied for its morphological accessibility and its importance to reflex action. Through the course of decades of research, scientists have documented the structure of the Mauthner cell system in detail. Some of the conclusions that have been drawn are that the Mauthner cell system differentiates between left and right source sound stimuli, encodes phase in an auditory context, and is responsible for mediating the threshold of the C-start response when reacting to visual stimuli. Certain studies have also shown that long-term potentiation of the inhibitory pathways of the Mauthner cell system contributes to the long-term plasticity of the system in responding to stimuli. Recently however, neurologists have discovered that the M-cell (see Glossary) system does not operate alone and is responsible for controlling many intuitive actions along with other neural circuits. Insight into how the reflex mechanisms of teleostei work as a result of the M-cell gives researchers an indication as to how similar reflex action circuits would function in humans. In this paper, I have proposed a new model that expands on older models to better account for the reaction of teleostei in a predatory C-start response.

Introduction

The study of the M-cell system helps neurologists to understand how decisions are made on the level of single cells. Neurologists have chosen to study the M-cell network due to its morphological accessibility and relative simplicity. Although I use the word "simplicity," the M-cell network is far from simple. As neurologists have discovered, the decisions controlled by this system require far more computation than was previously expected. They believe that if they understood the entire network in teleostei, they could extrapolate the results to higher order vertebrates and eventually studies could be conducted to reveal some of the processing networks in humans.

For those who are not familiar with some of the concepts being discussed in the paper, the following is a short background that explains some of the mechanisms by which signals are sent in neural networks.

An action potential, also referred to as a "spike", is the propagation of an electric wave down the axon (see Glossary) of a neuron. When the membrane is at resting potential (see Glossary), ion channels on the membrane surface remain shut, but as the potential rises to the threshold potential (see Glossary), the channels open, allowing sodium ions to flow into the axon from outside the cell. This event is called depolarization (see Glossary). This

results in a peak of the membrane potential, at which point the potassium ion-gated channels open and the potassium ions flow out from the axon. This returns the membrane potential to the resting potential as the sodium ion-gated channels close. Once the membrane potential reaches the resting potential, there is a short refractory period during which the potassium ions are pumped into the axon and the sodium ions are pumped out of the cell. An action potential is an "all-or-none" event, like the 1s and 0s in binary code, because if the potential does not cross the threshold potential, no action potential will be generated and each action potential has the same strength, regardless of the intensity of the stimulus. An example of an action potential is shown in Figure 1.

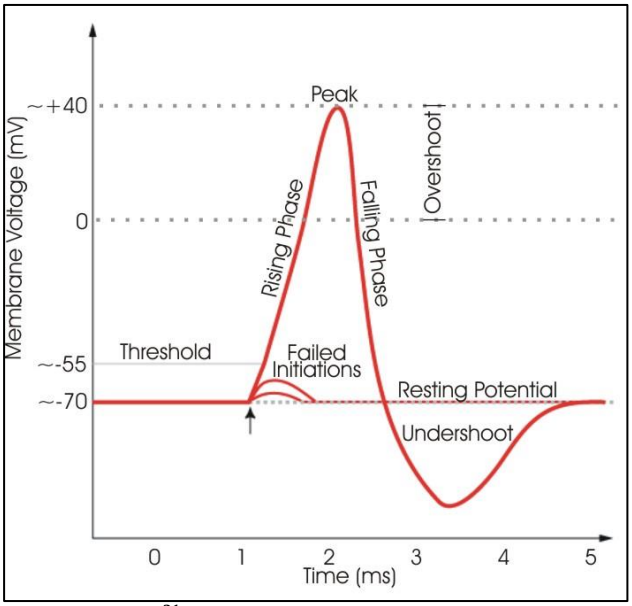


Figure 1*[31]*: *An example of an action potential.*

There are two forms of postsynaptic junctions: excitatory and inhibitory. When an action potential reaches an excitatory postsynaptic potential (EPSP), it induces the presynaptic neuron to release a chemical transmitter into the synapse, which in turn causes a depolarization in the membrane of the postsynaptic neuron. When an action potential reaches an inhibitory postsynaptic potential (IPSP), the chemical transmitter released by the presynaptic neuron causes a hyperpolarization (see Glossary) in the membrane of the postsynaptic neuron. An IPSP (see Glossary) can stop the flow of an action potential that originates elsewhere in a neuron, just as a dam can stop a flowing river. However, the size of an opposing EPSP can overcome the effect of an IPSP. The 'size' of an EPSP is determined by the frequency of the firing of presynaptic action potentials, which is positively correlated with the amount of the transmitter released at the synapse. Furthermore, according to the All or None Law of action potentials, the size of the action potential is independent of the size of the eliciting stimulus once threshold is exceeded. An example of a synapse (see Glossary) is shown in Figure 2.

[31]M. (n.d.). The generation of action potential in nerves. Retrieved July 14, 2016, from http://www.animalresearch.info/en/medical-advances/nobel-prizes/the-generation-of-action-potential-nerves/

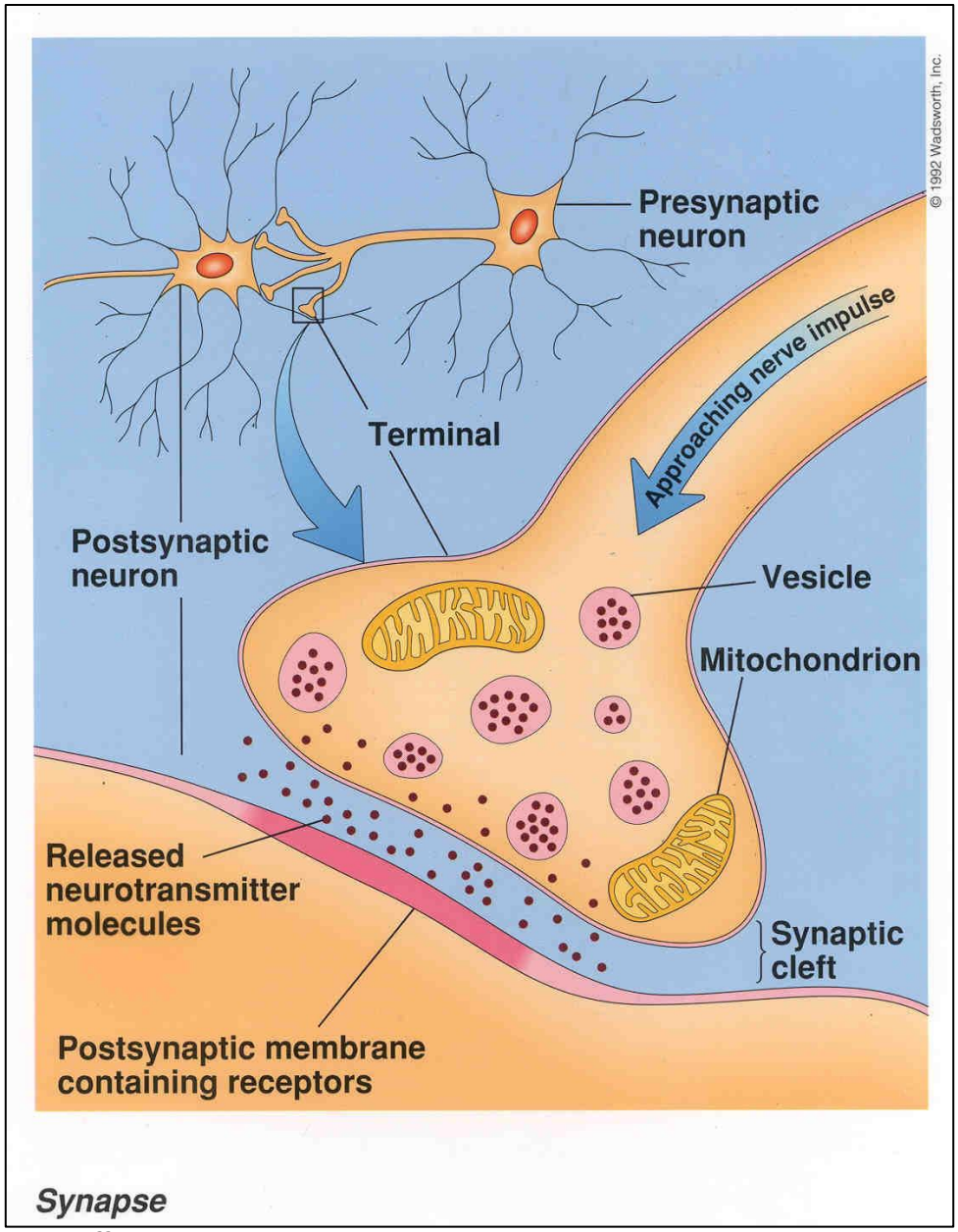


Figure 2[32]: *Image showing an axon synapsing onto a dendrite and neurotransmitters being released into the synaptic cleft*

[32] B., P. (2013, June 5). I-Search: Visual Aid: The Nervous System. Retrieved July 14, 2016, from https://prezi.com/5je2eahdig1i/i-search-visual-aid-the-nervous-system/

In teleostei (see Glossary), the Mauthner Cell System consists of two Mauthner cells with their axons crossing over each other. Each Mauthner cell, also known as an M-cell, has two major dendrites (see Glossary) and one axon running through the length of the animal's body. The axons of the M-cells cross over each other, so that each M-cell controls the side contralateral (see Glossary) to it.

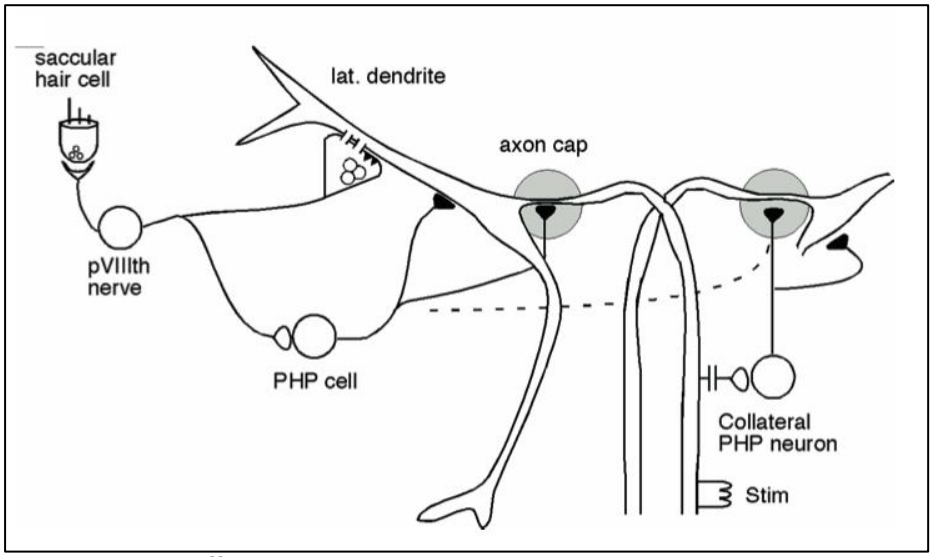


Figure 3[33]: *Diagrammatic representation of the left Mauthner cell*

The saccular hair cells (see Glossary) receive auditory stimuli and synapses on the pVIIIth nerve. This nerve has an axon that branches off into two. The first branch has an excitatory synapse on the lateral dendrite of the M-cell. The second branch has an excitatory synapse on a PHP neuron (see Glossary). This PHP neuron then has two inhibitory synapses on the M-cell, one on the lateral dendrite just before the cell soma (see Glossary), and one on the axon cap of the M-cell. From the axon of the M-cell, there is a smaller secondary axon that has an excitatory synapse on a collateral PHP neuron, which then has two inhibitory synapses on the second M-cell, one on the cell soma and the second one at the axon cap of the cell. Figure 3 illustrates this structure. Note that Figure 3 only shows one M-cell, but since the cells are symmetric, the structure shown for one M-cell also applies to the other.

The second dendrite of the M-cell, the ventral dendrite, receives information from the optic tectum (see Glossary), which, as the name suggests, receives information from the visual sensory organs of the animal. Like the pVIIIth nerve, the optic tectum, too, has two excitatory outputs, one to the ventral dendrite of the M-cell and the other to a PHP neuron, which has an inhibitory output on the soma of the M-cell. Figure 4 illustrates this structure.

[33] Weiss, S. A., Preuss, T., & Faber, D. S. (2009). Phase Encoding in the Mauthner System: Implications in Left-Right Sound Source Discrimination. Journal of Neuroscience, 29(11), 3431-3441. doi:10.1523/jneurosci.3383-08.2009

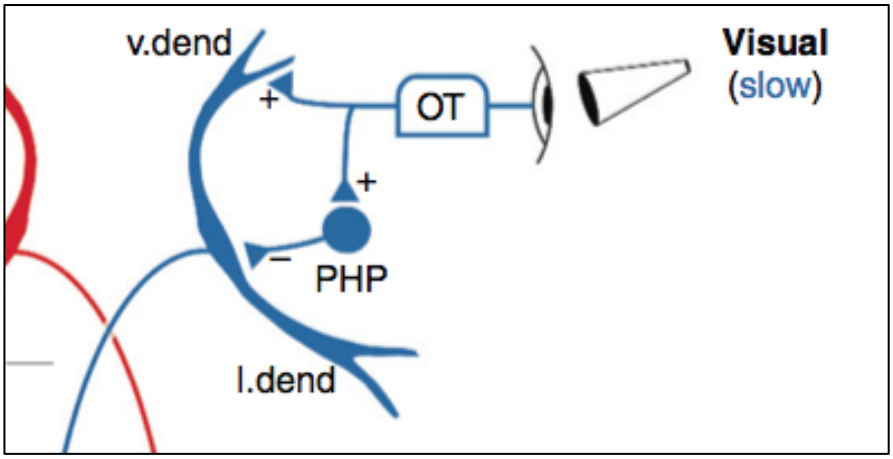


Figure 4[34]: *The blue neuron is the right Mauthner cell and the red neuron is the left Mauthner cell. Visual inputs to the Mauthner cell via the Optic Tectum (OT in the diagram). Only the right Mauthner cell has been shown. The left Mauthner cell would have the same inputs due to the symmetry of the cells. The axons of the M-cells cross over each other, as shown in the diagram.*

M-cells have been the object of study for scientists for three reasons:

1. They are very large compared to the other neurons, which makes them highly accessible

2. There are two cells, which makes matters less complicated when viewing the results of the experiments and

3. The action potential generated by the M-cell, also known as an M-spike (see Glossary), has a high intensity, which makes it easy to detect.

Researchers have explored M-cells in great detail, but this paper will only focus on two aspects of M-cells: their role in the C-start escape response to both auditory and visual stimuli.

Hypothesis

I hypothesize that Mauthner cells are responsible for the decision making process via the control of selective action patterns in teleostei, as well as for long term plasticity in the same.

Within the context of the program, it was not possible to conduct original experiments. Therefore, this paper sums up and analyzes the experiments conducted by other researchers, along with providing an example as to how we could gain deeper insight about our own neural networks as a result of these studies.

[34] Pfaff, D. W., Martin, E. M., & Faber, D. (2012, August). Origins of arousal: Roles for medullary reticular neurons. Trends in Neurosciences, 35(8), 468-476. doi:10.1016/j.tins.2012.04.008

Auditory Processing

The C-start evoked by an auditory stimulus is known for its short latency (see Glossary). This is because of the speed of the electrical synapses present between the pVIIIth nerve and the dendrites of the M-cell. The saccular hair cells receive an auditory stimulus. They generate action potentials, which travel along the pVIIIth nerve and go to a PHP neuron as well as onto the M-cell. If the number of potentials is large enough at the M-cell synapse, it can induce an action potential in the M-cell as well. An M-spike travelling through the axon stimulates the motor neurons, so the muscles contract and the fish travels in the direction away from the source of the sound. Figure 5 shows the C-start escape of a teleost to a stimulus.

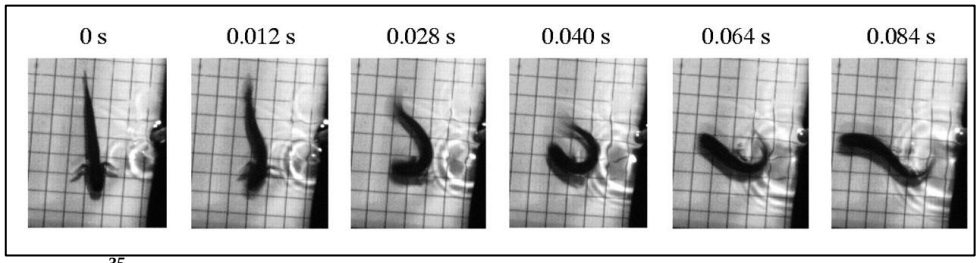


Figure 5[35]: *Series of images showing a C-start in a teleost. At 0.040s, the body of the fish turns into the characteristic C-shape, which gives the response its name. Note that the start is directed away from the stimulus source, which is indicated by the ripples in the water.*

A C-start escape is almost always directed away from the sound source, which leads to the question: How does the Mauthner cell discriminate between left and right sound sources? The discrimination between the azimuths (see Glossary) of the sound sources must be represented in the membrane potential changes of the M-cell. One of the most prevalent hypotheses seeking to explain this phenomenon is the phase model of directional hearing (Schuijf et al., 1975). According to this theory, the resolution of the azimuth of the sound source requires a phase comparison between the acceleration of the sound particle and the pressure of the sound (Eaton et al., 1995). The 'phase' of the sound wave is the angle (within 360) of the sinusoidal waveform at different points (Figure 6).

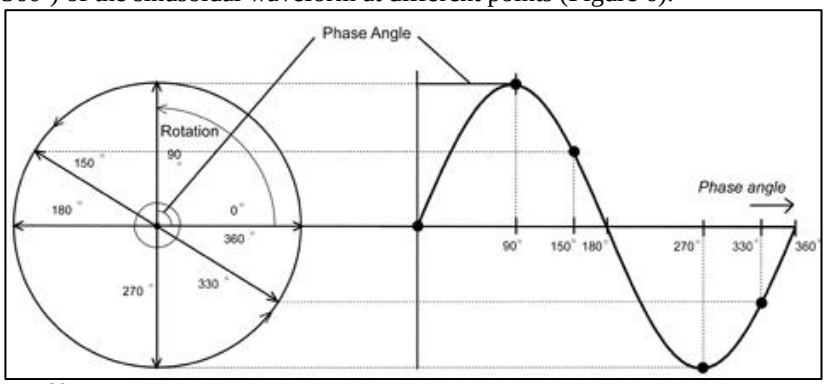


Figure 6[36]: *The calculation of the phase of a sound wave with the help of a polar plot.*

[35] Swanson, B. O., & Gibb, A. C. (2004, November 01). Kinematics of aquatic and terrestrial escape responses in mudskippers. *Journal of Experimental Biology, 207*(23), 4037-4044. doi:10.1242/jeb.01237

It was observed that the M-cell responded with a C-start response away from the sound source when the sound stimulus was initially compressed as well as rarefied (Canfield and Rose, 1996). In 1999, Eaton et al. proposed that this phase encoding is evident in the morphology, or anatomy, of the M-cell system itself. The team set out to find which network of the M-cell system best accounts for the balance between the inhibition and excitation that regulates the firing of the Mauthner cell and encodes phase at the same time.

They proposed that different endorgans (see Glossary) were responsible for encoding the two components of sound; the saccule (see Glossary) responded to the pressure component of sound and the utricle (see Glossary) responded to the motion of the particles. In the simulations they generated, there were four pools of hair cells in total, two pools on each side of the fish and each pool containing 16 units, or individual hair cells. On each side, there was one pool of hair cells in the saccule and one pool of hair cells in the utricle. The hair cells had excitatory afferents on the VIIIth nerve units. There were a total of four VIIIth nerve pools of ten units each, with two on each side, one receiving signals from the saccular hair cells and the other from the utricular hair cells.

The VIIIth nerve pools had excitatory afferents on the single Mauthner units on each side and on the two pools of PHP neurons on their side. The two pools of PHP neurons on each side had five units per pool. One of the pools, the bilateral PHP neurons, had inhibitory afferents on the Mauthner units on both sides and the pool of unilateral PHP neurons had inhibitory afferents on the Mauthner unit on the same side. This set up is shown in Figure 7.

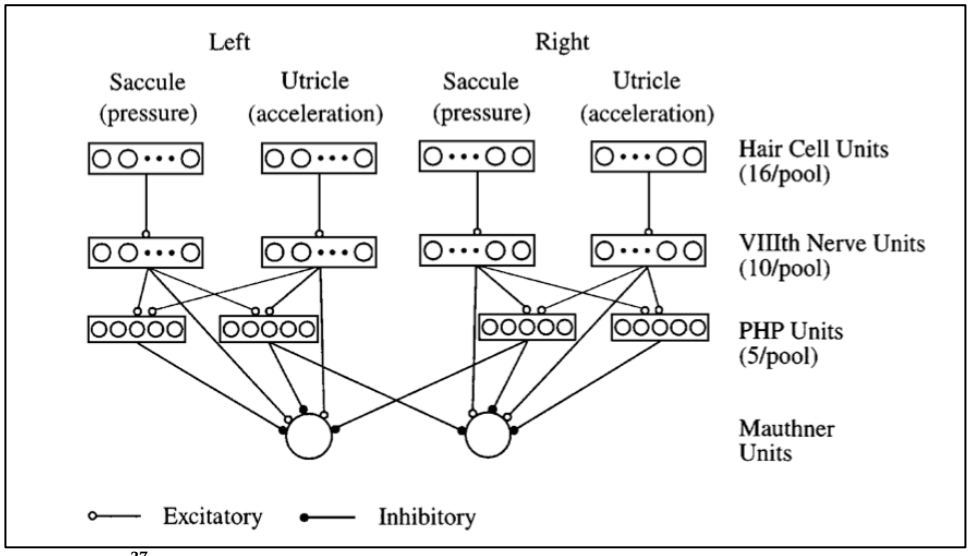


Figure 7[37]: *The arrangement of the pools and the levels in the Mauthner cell system.*

With the range of training stimuli presented to the networks, the team came up with 12 possible models for the M-cell system that accounted for the discrimination of the sound

[36] The Physical Principles of Sound. (n.d.). Retrieved July 11, 2016, from http://www.jiscdigitalmedia.ac.uk/guide/the-physical-principles-of-sound

[37] Guzik, A. L., Eaton, R. C., & Mathis, D. W. (1999, March). A Connectionist Model of Left-Right Sound Discrimination by the Mauthner System (G. Laurent, Ed.). *Journal of Computational Neuroscience, 6*(2), 121-144. doi:10.1023/A:1008828501676

source according to the phase model of discrimination. In order to identify which type of network was most accurate, the team compared its results to those generated by Faber and Korn in 1978 (Figure 8).

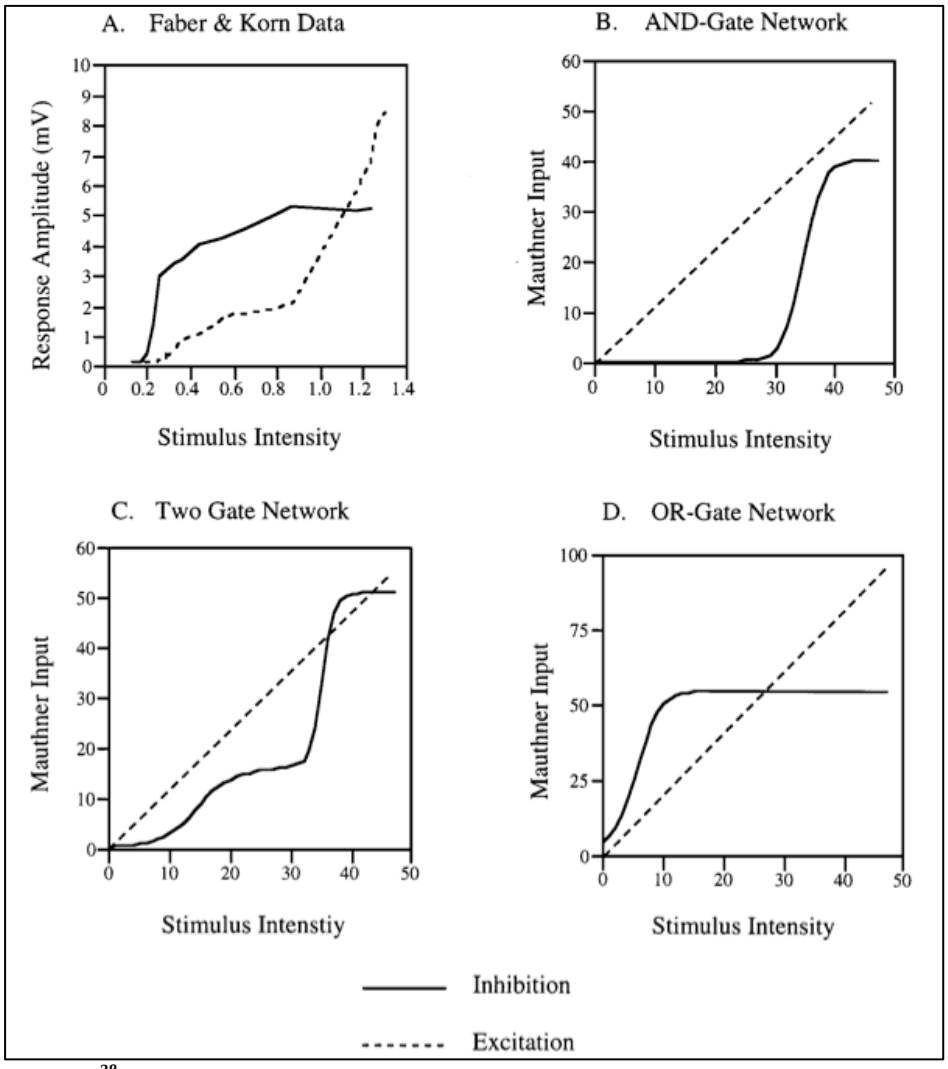


Figure 8[38]: *Comparison of Faber and Korn's data (A) and Eaton et al.'s data (B, C, D). The data generated by Eaton's team for the OR-gate configuration (D) most closely matches the data generated by Faber and Korn (A).*

Faber and Korn measured the difference between the excitation and inhibition of the M-cell system by sending electric shocks of different intensities to the VIIIth nerve and the

[38, 14] Guzik, A. L., Eaton, R. C., & Mathis, D. W. (1999, March). A Connectionist Model of Left-Right Sound Discrimination by the Mauthner System (G. Laurent, Ed.). *Journal of Computational Neuroscience, 6*(2), 121-144. doi:10.1023/A:1008828501676

PHP neurons. Eaton's team had constructed formulae that modeled the activity of the neurons in their networks. Once they ran these formulae, the data generated indicated that the networks that best fit the experimental data were those in which the PHP neurons functioned as OR-gates. Figure 9 shows an example of an OR-gate network.

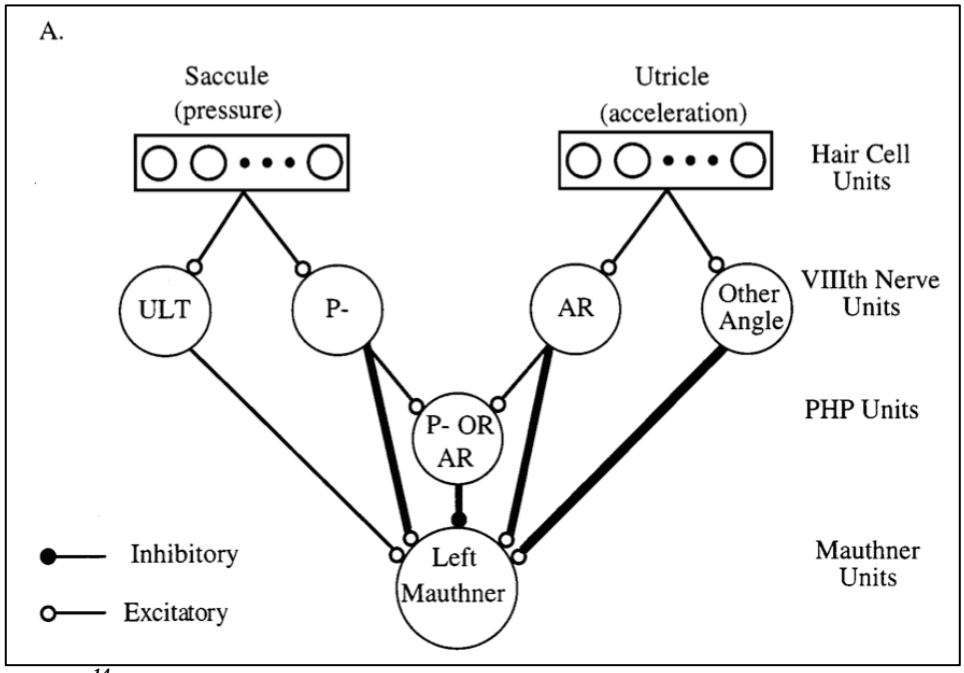


Figure 9[14]: *An example of an OR-gate configuration devised by Eaton et al. to account for the processing of the auditory stimulus in the M-cell system.*

The team concluded that the networks would only function if the Mauthner cells had high threshold potentials. Otherwise small excitations from the rest of the synapses would be enough to fire the M-spike. Furthermore, they also inferred that only PHP neurons with high threshold potentials would be able to solve the logical computations, whereas the PHP neurons with low threshold potentials would only be able to solve amplitude discrimination. This is because PHP neurons with low thresholds would respond to the smallest excitations, so even if one input were strong, the absence of the second input would still result in the activation of the PHP neuron, which would be detrimental. This experiment was carried out under the assumption that the Mauthner cell itself did not play a large role in the processing of information. Therefore, the team simply investigated the computation in the inhibitory circuits of the M-cell system.

In 2009, Weiss et al. conducted an experiment in order to test the phase model for the left-right source discrimination of sound underwater.

> *"A critical component of the phase model is that phase information be preserved in the responses of the M-cell system" (Weiss et al., 2009).*

One of the ways in which phase information is preserved through the responses of the M-cell system is if the firing of the action potentials in both the axon of the M-cell and in the feed forward inhibitory neurons is phase locked to the sound wave. This is made possible by the electric synapses in the M-cell system, which are responsible for the short

latencies of the response from stimulus onset. Weiss postulated that the integration of both the excitatory and inhibitory inputs in the M-cell system might help discriminate left-right sound sources in the Mauthner cell.

In their experiment, Weiss et al. used ramped sinusoidal sounds of three frequencies: 150, 200 and 250 Hz. In order to record the changes in the membrane potential, EMG (see Glossary) electrodes were placed within the cell. High-speed video cameras were also used to record the physiological response of the fish to the stimuli.

To mathematically analyse the phase angle of the sound stimulus, the analytic signal was then converted to a polar form that had been devised by Smith et al. in 2002. The mean phase angles and the primary and secondary distributions, as well as the vector strengths of the same, were analysed using a polar coordinate formula developed by Fisher (et al.) in 1993. To correctly estimate the phase angles that evoked the responses in the M-cell system, the team also adjusted the angles to account for the sensory processing time, from the stimulus onset to the actual response in the M-cell. A further 1.5 ms was subtracted to account for the EMG latency and for the short delay between the axon of the M-cell and the neuromuscular junctions. The value of 1.5 ms was taken from the results of the experiments conducted by Zottoli in 1977. Once the phase models for each of the trials was analyzed, the team concluded that the more coherent cluster with the larger fast EPSPs (see Glossary) was the primary distribution, and the less coherent cluster with the smaller events was the secondary distribution.

Weiss et al. observed that fast EPSPs were set off as a result of the stimulus and were superimposed on the depolarizing envelope of the slow EPSP, which gave rise to a pattern in which PSPs were observed to be increasing in intensity as the intensity of the sound stimulus increased. An example is shown in Figure 10.

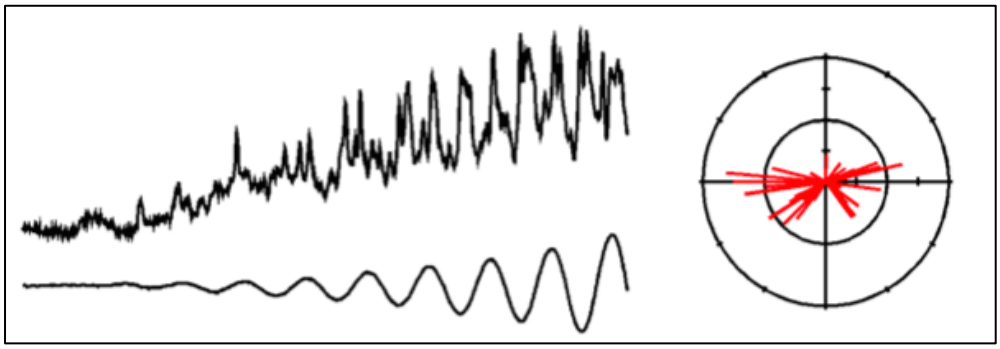


Figure 10[39]: *EPSPs generated in the M-cell in response to the 200Hz stimulus. On the left, two graphs have been given. The lower graph denotes the sound stimulus, which increases in intensity with time. The upper graph indicates the responses recorded within the M-cell to the stimuli. Note that on the right side, there is a polar plot that shows the different angles, or phases, of the sound wave to which the M-cell responded (with peaks). There are two major distributions, with the primary distribution falling along the border of the second and third quadrant and the secondary distribution falling along the border of the first and fourth quadrant. The length of the lines indicate the vector strength of the response.*

[39] Weiss, S. A., Preuss, T., & Faber, D. S. (2009). Phase Encoding in the Mauthner System: Implications in Left-Right Sound Source Discrimination. Journal of Neuroscience, 29(11), 3431-3441. doi:10.1523/jneurosci.3383-08.2009

A peak detection program detected the peaks of the fast EPSPs as well as their phase angles relative to that of the sound wave. The team discovered that the Mauthner cell system did, in fact, encode phase and that the primary and secondary distributions were locked to diametrically opposite angles. The team concluded that the vector strengths did not depend on the frequency of the sound wave, but that the phase angles depended on the frequency of the sound stimulus.

The fact that the primary and secondary distributions were locked to diametrically opposite phase angles of the sound wave indicates that the M-cell system fired to both compression as well as rarefaction. The team also checked whether the primary and secondary distributions changed if the stimulus intensity was increased. However, the distribution angles remained the same, which indicates that the sensory processing time is independent of the intensity of the stimulus.

The finding is quite significant, for it indicates that once the sound stimulus passes a certain intensity level, the M-cell system fires and evokes the C-start response. Each escape becomes vital at that point, irrespective of the intensity of the stimulus. This form of firing may provide the fish an evolutionary advantage, for it means that the M-cell system would fire with the same speed for big predators as well as small, enabling the teleostei to escape from a vast variety of dangers, aiding their survival.

The team also checked if the phase encoding would be apparent at the behavioural level of the fish. Since EMGs are found simultaneously all along the rostral-caudal (see Glossary) axis of the fish at the beginning of the C-start response, they are a good indicator of the timing of the M-spike and, thus, of the latency of the response.

Once again, the team found that the timing of the M-spike was phase locked to the sound stimulus. Furthermore, once the sensory processing timings and the instrument recording latencies were accounted for, it was observed that the M-spike was evoked by the "increasing and decreasing phases of sound compression" underwater (Weiss et al., 2009). Another important finding was that the responses did not occur once the sound intensity level reached a set threshold, but that the response was elicited over a range of intensities. This is counter-intuitive, as it is expected that the fish would respond only once they actually hear the sound, which would be a set threshold limit for most fish. A possible evolutionary advantage this confers on the fish is that it takes the predators by surprise when the fish do not respond to a certain noise level one time, but respond to it at a later time. This could aid the fish in escaping the predator by taking advantage of its confusion.

Weiss et al. next determined whether the activity in the inhibitory PHP neurons was also phase locked to the sound wave. They recorded from outside and inside the neuron and computed the difference between the two recordings as the transmembrane (see Glossary) potential. They discovered that unlike the M-cell, the PHP neurons had mostly suprathreshold (see Glossary) responses for the ramped sound stimulus, which were phase locked as well.

The team then checked the effect of the inhibition on the activity of the M-cell. They recorded intracellularly from the Mauthner cell soma and extracellularly from the axon cap of the Mauthner cell (Figure 11). They determined the transmembrane potential by subtracting the extracellular recording from the intracellular recording. It was observed that the transmembrane potential did not deviate much from the intracellular recording because the size of the EHP (see Glossary), or the extracellular recording, was so small. The EHP increased in magnitude only for later cycles of the stimulus.

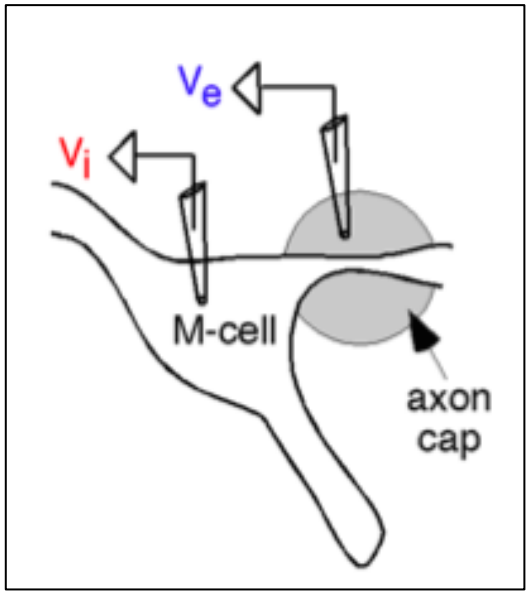


Figure 11[40]: *How the intracellular and extracellular recordings were taken. The shaded circular area is the axon cap, which is the region where the EHPs can be recorded. V_i is the intracellular recording electrode and V_e is the extracellular recording electrode.*

This finding implies that for the initial stages of the sound stimulus, there is very little inhibition on the Mauthner cell. This is important for the fish, as the Mauthner cell needs to respond to the sound stimulus with a C-start escape, following which inhibition is required to stop the M-cell firing again; otherwise the fish would keep turning. If the Mauthner cell fired more than once, the sheer magnitude of the M-spike would damage the fish and the fish would keep turning on the spot without actually escaping.

The team also tested the response with a pip stimulus, which is an abrupt loud sound. The result was that electrical inhibition was much stronger as the firing of the PHP neurons was synchronised. This mechanism is important to ensure that the fish does not respond with a C-start escape for every small disruption that occurs. Multiple M-spikes, as mentioned earlier, would prove detrimental to the fish, as the size of the action potential places great stress on the muscles of the fish's mid-body.

To investigate the role that inhibition plays in a C-start escape, the team added strychnine to the PHP neurons in small amounts. Strychnine inhibits the release of the neurotransmitter (see Glossary) glycine, which is present at inhibitory synapses. The result was that the inhibition was significantly lowered and the M-cell fired earlier than the control trials before strychnine was added. However, the addition of strychnine did not affect the sensory processing time or the phase locking of the M-spike. It simply reduced the size of the PHP, resulting in an increase in the vector strength of the M-spike.

The role of long-term potentiation in the plasticity of the M-cell system has also been extensively researched. Indeed, researchers have discovered that long-term potentiation of

[40] Weiss, S. A., Preuss, T., & Faber, D. S. (2009). Phase Encoding in the Mauthner System: Implications in Left-Right Sound Source Discrimination. Journal of Neuroscience, 29(11), 3431-3441. doi:10.1523/jneurosci.3383-08.2009

the inhibitory pathways of the M-cell system plays a large role in modifying its response to stimuli.

> *"Long-term potentiation (LTP), the increase in synaptic strength evoked by high-frequency stimulation, is often considered to be a cellular model for learning and memory."*
>
> *Oda et al., 1998*

Oda and his team studied the behavioral effect of long-term potentiation of the inhibitory pathways of the M-cell system in 1998. Since the M-cell response system is regulated by the inhibitory action of the feed forward inhibitory neurons, Oda et al. tested long-term potentiation and its physiological effects on the C-start escape of goldfish. However, it was not just Oda who tested the effect of long-term potentiation on inhibitory pathways. Korn et al. in 1992 and Charpier et al. in 1995 also explored the effects of long-term potentiation. Due to limited space, only the work by Oda et al. in 1998 will be expanded on in this paper.

The team used conditioning tone bursts of 500Hz that were 50ms long every four seconds for three to five minutes from an underwater loudspeaker. They evoked escape responses by dropping a 3.3 cm wide plastic ball that weighed 18g into the water every minute for five trials and then taking a break for five minutes between sets of trials.

The experiments concluded that there was long-term potentiation (LTP) in the inhibitory pathways in the M-cell system, which lasted over five hours. It was observed that both electrical as well as auditory conditioning, which led to LTP, enhanced the inhibition in the M-cell system without altering the strength of the EPSPs, leading to a net increase in inhibition.

Finally, when the team actually compared the difference in the behavioural responses of the fish before and after auditory conditioning, they found that the number of responses to the stimuli decreased by over 70% in 9 of the 12 fish tested. Furthermore, the time taken for the fish to revert to responding to the stimuli at preconditioning levels was approximately two hours. However, the direction and the latency of the responses remained constant, even though the number of responses was lower. This indicates that the auditory conditioning only enhanced the inhibitory pathways in the M-cell system and did not affect the sensory processing times or the timing of the fast EPSPs. As expected, the frequencies that the fish best responded to and that were used for the auditory conditioning had the largest decrease in response probability.

The experiment showed that the LTP of the inhibitory pathways enhanced the inhibition but left the excitatory pathways in the M-cell system untouched, and this eventually led to a behavioural modification as well. This mechanism, whereby inhibitory LTP can alter the reflex action of a fish, might serve as an evolutionary advantage for fish in turbulent or noisy waters, where responding to every stimulus might become detrimental for the fish. Instead, the organism gets attuned to sounds of similar frequencies as a result of LTP and stops responding to every stimulus of that frequency.

Visual Processing

The second half of this paper focuses on the role the M-cell system plays in the processing of visual stimulus and responding to the same with a C-start escape. Unlike C-start escapes in response to auditory stimuli, which were based on the need for prey to move away from their predators, C-starts initiated by the M-cell system in response to visual stimuli can be directed both towards and away from the source. Furthermore, the latency for the responses is two orders of magnitude more than the latencies for responses to auditory

stimuli. This points to a more complicated neural network that is involved in a complex decision-making process and one that might not be purely a reflex.

First, we consider the instance when the C-start escape evoked by visual stimuli is to escape the predator. The mechanisms by which the sensory stimuli are transformed into a motor command remain mostly elusive (Preuss et al., 2009). Preuss and his team undertook research in 2009 to try to predict a function that relates the stimulus to the escape probability as well as its timing. In order to investigate the relationship between object size and object acceleration to the response timing of a fish, Preuss et al. used eight different stimuli. Some had the same initial size with differing approach velocities, while others had the same approach velocities with differing initial sizes (Figure 12).

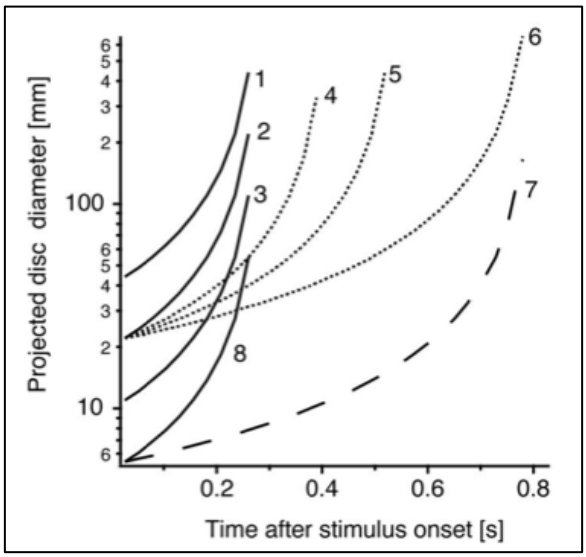


Figure 12[41]: *The different disc stimuli that were used in the experiment. Stimuli 1, 2, 3 and 8 have the same approach velocity. Stimuli 2, 4, 5 and 6 have the same initial sizes but differing approach velocities. Stimulus 7 is an extreme stimulus with a small initial size and a slow approach velocity.*

They found that when the fish was in the direct collision trajectory of the object, it showed no preference for left or right turns. This uncertainty in the response is expected as it helps the fish to remain unpredictable, and by reacting as late as possible, it would be impossible for the predator to change its course. They also found that larger and faster stimuli had shorter escape latencies, which means that the C-start was evoked earlier than for small, slow stimuli. This is also expected because a bigger object usually means a more dangerous object, so the fish would naturally get out of the way much earlier.

Another finding was that the fish did not react to a black disk that decreased in size, or seemed to move away from the fish. Again, it is quite logical that this would be so, as there would be no danger from a receding object. However, a surprising discovery was that none of the fish reacted in any trials when a black disk of fixed size suddenly appeared on the

[41] Preuss, T., Osei-Bonsu, P. E., Weiss, S. A., Wang C., & Faber, D. S. (2006, March 29). Neural Representation of Object Approach in a Decision-Making Motor Circuit. *Journal of Neuroscience, 26*(13), 3454-3464. doi:10.1523/jneurosci.5259-05.2006

screen. It is unclear what advantage this could possibly have for the fish. Perhaps, it is possible that the object may be perceived as prey, not a predator, and since the disk did not increase in size, it apparently did not pose any danger to the fish, so there was no need to react. It could also be that the lack of response to the sudden appearance of an object indicates an imperfect circuit. After all, evolution may not have accounted for a response to this form of stimulus for multiple reasons:

1. There was not enough selection pressure with this form of stimulus to attune a neural circuit to respond to it.
2. Evolution is dynamic and the neural circuits of the fish still have to adapt to this form of stimulus.

With the eight stimuli, the team had tested the responses of the fish to approach speed, time to collision and angular retinal image size of the object. Now, they set out to determine which of the three conditions, or parameters, was responsible for evoking a response in the fish. They reasoned that if any one variable were responsible, then the variable would have a constant value for all trials in all fish at the response onset, accounting for sensory processing latency.

The results showed that although the behavioral threshold was seen to correlate more with approach speed and angular retinal image size than time to collision, no single parameter was responsible for evoking a response in the fish. However, the team noticed that the behaviour occurred before the object finished its approach. They ruled out the role of input saturation of the retinal image for the following reasons:

1. The mean angular size of the image when the response occurred was 21 °, whereas the visual field of the goldfish eye is 190 °.
2. Some looming sensitive neurons in zebrafish have receptive fields up to 30 °.

The team noticed that the stimulus became less effective just before it reached its maximum angular size. Therefore, they tested another function that peaked before stimulus offset to check if that explained the behaviour of the fish. The function, $\eta(t)$, was based on two parameters, which are image expansion velocity and retinal image size. They also tested two additional formulae, termed $\kappa(t)$, which was based on angular image size alone, and $\omega(t)$, which was based on image expansion velocity alone.

Their results suggested that $\kappa(t)$ was the function that best fit the requirement of a constant value at the response onset, and therefore was probably represented in the M-cell system during object approach. The function predicts that the PSP (post synaptic potential) peaks early and then decays while the stimulus continues to approach, so the function should match the M-cell recordings.

The team discovered that the latency of the PSP peak time to the stimulus onset was smallest for larger and faster stimuli, and the peak times predicted by the Kappa function were in close agreement with the experimental results. However, the functions $\eta(t)$ and $\omega(t)$ also showed close agreement with the experimental results. The team therefore deduced that the PSPs trigger the M-cell when it is at or close to its peak depolarization.

To understand why the PSPs peaked before the object finished its approach, the team tested the PHP cells to determine the role of feed forward inhibition. They recorded EHPs that occur at the axon cap of the M-cell and the potential within the cell body. The results indicate that after the PSPs peak in the M-cell, the IPSPs and IPSCs (see Glossary) increase, which reduces the peaks of the PSPs. This inhibition plays an important role in stopping the EPSP, which would otherwise result in more action potentials being sent down the M-cell axon, causing the fish to keep turning. This is quite similar to the role of the PHP cells in the response of the M-cell system to auditory stimuli.

The study showed that although the retina and optic tectum first receive and process the visual stimulus, the M-cell system is the deciding factor in when the C-start escape is evoked. The inhibitory pathways in the M-cell system play a vital role in controlling the magnitude of the EPSPs, and are consequently responsible for determining the threshold of the M-cell and the timing of the escape. This was also the first study that proved the role of the Mauthner cell in a response to a stimulus that was not simply abrupt but smooth and gradual. In fact, as the researchers saw, the fish did not respond to abrupt stimuli at all in the experiments.

The team also concluded that the fish does not respond purely to angular retinal image size or the image expansion velocity and instead found that the function that best describes the response of the fish is scaling function $\kappa(t)$. The use of a function to describe different aspects to which the fish respond indicates that these variables, or parameters, are inherently encoded in the neurons that are afferent on the ventral dendrite of the M-cell.

In the second half of this section, we consider the instance when the C-start is evoked by the M-cell system, so that the fish can move in the direction of the prey to capture it. This form of reaction is seen predominantly in archerfish. Archerfish are surface hunters, and so they face a lot of competition in terms of catching their prey due to the large presence of other predatory fish (Schuster, 2011). Once they shoot down their prey with a jet of water, they have to navigate and reach the place where the prey is projected to fall so that no other fish can get the prey. They time their distance and speed of push off at the end of the C-start to arrive at their target the moment it hits the water. Evidence has been presented to show that a particular neural circuit can control more than one action. Therefore, researchers have inferred that the predictive starts of the archerfish are controlled by the M-cell system. However, it is not purely a reflex action, as there are approximately 1600 fine motor adjustments that are made before the fish reaches its prey's landing point.

The three main variables that the fish have to take into account are the height of the prey above the water, its direction of fall and the speed at which it travels (Schlegel and Schuster, 2008). In their experiment, Schlegel and Schuster discovered that the latencies and accuracy of the archerfish did not change when the prey was released without providing the fish with shooting-related contextual clues ("invisible" prey), compared to when the fish shot the prey and reacted to it falling. This could possibly indicate that the shooting of the prey is not vital to the response of the fish, which relies purely on the variables mentioned earlier.

The scientists investigated further and set up an experiment in which there were three possible release sites. They would direct the fish to any one of the three release sites, but release the "invisible" prey randomly from any of the three release sites. The results showed that the latency was not affected even for a distance of 20cm and that increases in the latency were apparent when the prey was 40cm away from the fish. Furthermore, the accuracy of the fish was completely unaffected in all the experiments, no matter how far away it was from the prey. Figure 13 shows the set up of the experiment, with the contrast between the normal manner by which archerfish acquire their food and the "invisible" prey.

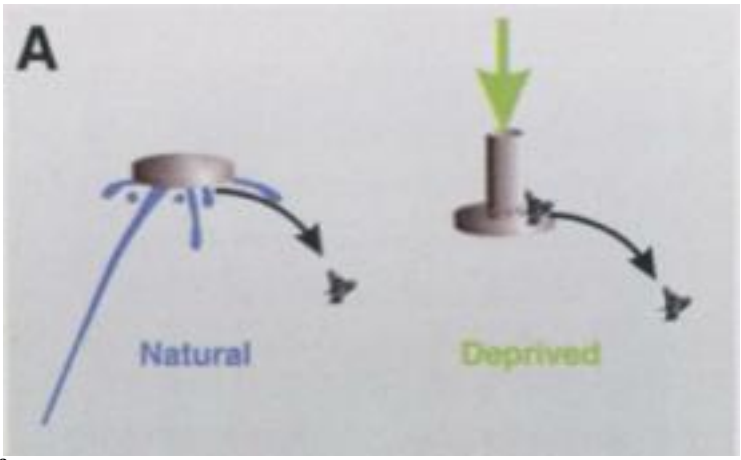


Figure 13[42]: *Left shows how an archerfish normally shoots its prey and then responds to its motion after the shot. Right shows the experimental set up of the "invisible" prey. The prey were inserted into the top of the release point and then shot through the side at a random velocity and direction. This set up ensured that the prey were "invisible" to the fish before release.*

To determine if the fish averaged motion signals over large regions, the researchers simultaneously released two "invisible" prey from the same point but in different directions (Figure 14). They discovered that the latency times and accuracy for the fish remained unchanged, and that the fish would always select one of the two prey and propel itself to the landing point of that prey. Furthermore, the decision of choosing which prey to pursue was not made at random. The scientists observed that the fish always went for the prey that would fall closer to it.

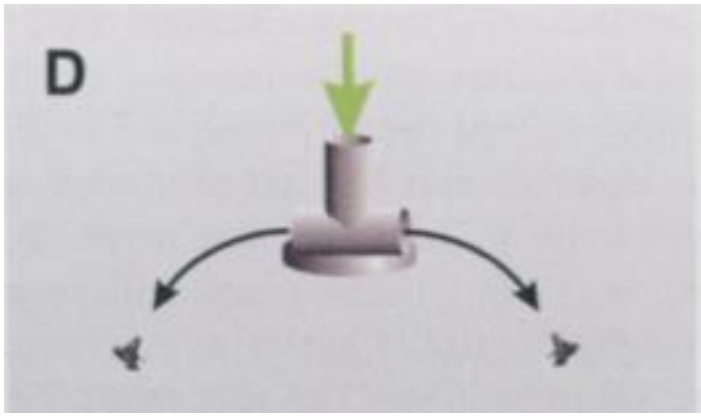


Figure 14[43]: *Set up of the experiment in which two prey were released simultaneously from the same release site.*

[42] Schlegel, T., & Schuster, S. (2008, January 04). Small Circuits for Large Tasks: High-Speed Decision-Making in Archerfish. *Science*,*319*(5859), 104-106. doi:10.1126/science.1149265
[43] Schlegel, T., & Schuster, S. (2008, January 04). Small Circuits for Large Tasks: High-Speed Decision-Making in Archerfish. *Science*,*319*(5859), 104-106. doi:10.1126/science.1149265

In the process of testing more variables, they found that although there were occasional changes in the latency, there were no changes in the accuracy. This could be attributed to the fact that the fish would rather have a slower reaction time than miss the target altogether, as that would result in more taxing course corrections and the fish might even miss the meal due to competition. The lowest latency that the team was able to obtain was 40ms. Since it is estimated that there is a delay of 35ms (Zottoli et al., 1987) in the visual processing circuit of goldfish, the fact can be extrapolated to estimate that the processing time in an archerfish is not too different. If the information processing time is capped at 10ms, it still implies that the computation must be taking place within the retina itself, and the signal follows the fastest neural pathway to the M-cell system. According to them, the computation of the fish's spatial position and accommodation takes place in the "latter part of the circuitry" before the response is carried out.

Schlegel and Schuster's study shows that even highly complex decisions can be carried out in split seconds by a simple neural circuit. This holds important implications for the way we view the complexity of neural circuits in relation to the complexity of the action they control. This idea will be further expanded in the discussion section of the paper.

Wöhl worked with Schuster to investigate the kinematics of the archerfish's predictive start. The data they obtained regarding the linear velocity and acceleration and angular velocity and acceleration of a predictive start closely mirror the data for a C-start. In 1983, Bellman and Krasne's experiment provided evidence for the fact that the same neural circuit could be used to control two different actions altogether. Since predictive starts are incredibly quick, they must be controlled by relatively small neural networks of reticulospinal neurons. Canfield and Rose's experiment in 1993 provided evidence for the involvement of the M-cell system prey capture for goldfish. However, they reasoned that the predictive start does not depend entirely on the M-cell network but also on another pair of large cells called MiD2cm and MiD3cm. In conclusion, the role of the M-cell system in initiating the C-start at the beginning of the predictive start is undeniable.

Conclusions

The results of the experiments led to the following conclusions:

1. Electrical synapses play a vital role in maintaining phase information of the sound stimulus by reducing the sensory processing

2. M-cell system encodes phase

3. Long term potentiation of inhibitory pathways plays a role in the plasticity of the M-cell system in response to auditory stimuli

4. In terms of visual processing, the Mauthner cell system plays an important role in both escape as well as prey capture decision making

5. The longer latencies of the response to visual stimuli indicate that the action is not purely a reflex but a voluntary action with a much more complex neural network at play than the network for auditory processing.

6. The M-cell system is responsible for the long-term plasticity of the responses to visual stimuli as well, for it was shown that fish usually responded to prey with predictive starts that had been rewarded in the past.

Discussion

The Mauthner cell system is easily accessible in fish and also relatively simple. Yet, as we have seen in this paper, it controls a vast number of actions in the fish, most of which are crucial to its survival. The manner in which the M-cell system encodes the different stimuli in its neural networks could help neuroscientists understand how information is transmitted in the form of chemical signals.

The studies explored in this paper have shown that the M-cell system is amazingly dynamic and is capable of altering itself to provide long-term plasticity in its response to visual and auditory stimuli. This is due to the long-term potentiation (LTP) of the inhibitory pathways in the neural circuit. This indicates that it is not the excitatory synapses but the inhibitory synapses that mediate the threshold of the response, and therefore it is where the computation of different aspects of the stimuli takes place.

Furthermore, researchers have found that even when the Mauthner cells were lesioned (see Glossary) in goldfish, they were still able to perform a strong C-start escape. This clearly points to the fact that neural circuits that mimic the M-cell system play a role in the C-start. They cannot support the M-cell in the response, as that would imply that they are unable to execute the C-starts themselves. Sure enough, Fetcho et al. (1996) showed that two homologs to the M-cells called MiD2cm and MiD3cm fire along with the M-cells when the stimulus is rostral to the fish. This occurs so that the fish can turn almost 180 ^{o}in order to escape from the stimulus.

An interesting observation that has come up as a result of the studies being conducted into the decision making process in teleostei is the complexity and speed of the archerfish's predictive start. This discovery leads to intriguing questions regarding the role of the neural circuits at play in processing the information necessary to execute the predictive start. Could the MiD2cm and MiD3cm also be playing a role in this action?

I certainly believe that the Mauthner cell does not act alone in the predictive start. Researchers have discovered that even when the Mauthner cells were lesioned, the archerfish can still execute the predictive starts, albeit with slower latencies. The neurons in the hindbrain and parts of the spinal cord of zebrafish have been mapped out. Since most of the neurons in lower order vertebrates are similar to each other, I will be using the data generated from studying the zebrafish's neural systems to extrapolate a possible neural network that accounts for the processing involved in generating the predictive start as well as the C-start escape response.

As mentioned earlier, the latency of the response compared with the complexity of the computation required for the predictive start indicates that the neural network involved in the response is relatively simple, and inhibitory neurons and collaterals of the axons in the neurons are responsible for the processing of the information. For the signals to reach the motor neurons in the trunk of the fish within a span of 22ms, the stimulus must be processed in a minimum number of cells, which have relatively large action potentials in order to sufficiently cause the body of the fish to turn in the direction of the prey. Further motor adjustments would be made by the smaller neurons in the spinal cord and the hindbrain of the fish. This assumption necessitates that these smaller neurons also receive input directly from the sensory organs but have slightly more complex processing networks. They might also have higher thresholds than the larger neurons, like capacitors in a time delay circuit, which would make them fire a little later. This would account for the delayed response of these motor neurons in correcting the course of the fish. Furthermore, the response would not depend on just one impulse from the retina of the fish but constant inputs, which would provide information updating the fish about its spatial orientation and objects in its path.

Therefore, it can be inferred that because the fish has to avoid other fish or objects in its path, there is more than one neural network at work in controlling this response.

The following figure is a possible neural network that seeks to explain the escape response of the fish. Unlike Eaton et al. (1999), this network also includes the MiD2cm and MiD3cm cells following further research, which implicates the role of these homologs in the C-start escape. However, I will use Eaton's original network as a foundation and I will expand on it to include the MiD2cm and MiD3cm cells in the network.

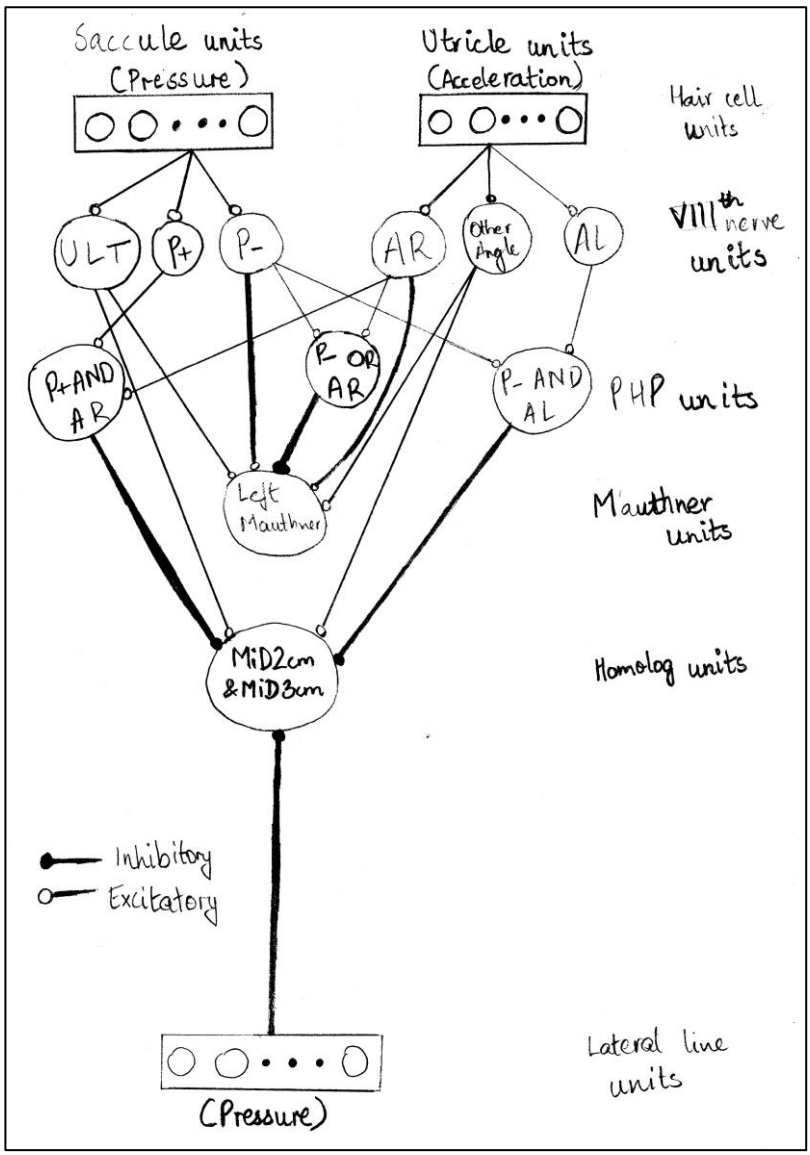


Figure 15: *A possible neural network that includes the MiD2cm and MiD3cm cells along with the Mauthner cells in the escape response. The thickness of the lines indicates the weight that the certain neuron carries in terms of excitation or inhibition. There were ten hair cell units in the saccule, utricle and lateral line each to sense the auditory stimulus.*

As described by Eaton et al. (1999), the ON configurations for the left Mauthner cell are as follows:

1. Pressure is negative (P-) and acceleration of the particles is from the right to the left (AR)
2. Pressure is positive (P+) and acceleration of the particles is from the right to the left (AL).

It goes without saying that the OFF configurations will be the opposite of the aforementioned arrangements.

The functioning of the network for the Mauthner cell has already been documented by Eaton et al. (1999). I will expand on how the circuit affects the functioning of the same circuit but after including the M-cell homologs. The functioning of the homolog cells depends primarily on the time difference between the inhibition from the lateral line hair cell units and the excitation from the saccule and utricle hair cell units. It is for this reason that the lateral line hair cell units are positioned as far caudally as possible. These cells have strong inhibitory afferents on the homolog cells. The synapses are also close to the axon of the homolog cells, as this would provide the most effective inhibition. When the stimulus is rostral to the fish, the ULT units and Other Angles units have excitatory inputs to the neuron much before the inhibition from the lateral line units arrives, as the stimulus would be processed in the head quicker than near the tail. In this case, the homologs fire, supplementing the action of the M-cell. To ensure that the homologs do not fire to any of the OFF combinations for the ipsilateral (see Glossary) M-cell, there are two AND-gate PHP units that code for the OFF combinations and provide strong inhibition to the homolog cells. When the stimulus is caudal to the fish, then it will activate the lateral line units before the saccule and utricle units. These lateral line units will then inhibit the homologs from firing as the inhibition will hyperpolarise the membranes of the homologs. This network accounts for the documented observations and is also devoid of complexity, which ensures that its latency matches that of the M-cell system.

The neural network that handles the response to visual stimuli is much more complicated as it takes into account various factors such as height, initial speed and direction of the prey. Also, there are many more neurons responding to the same stimulus in order to ensure a high level of precision in the action.

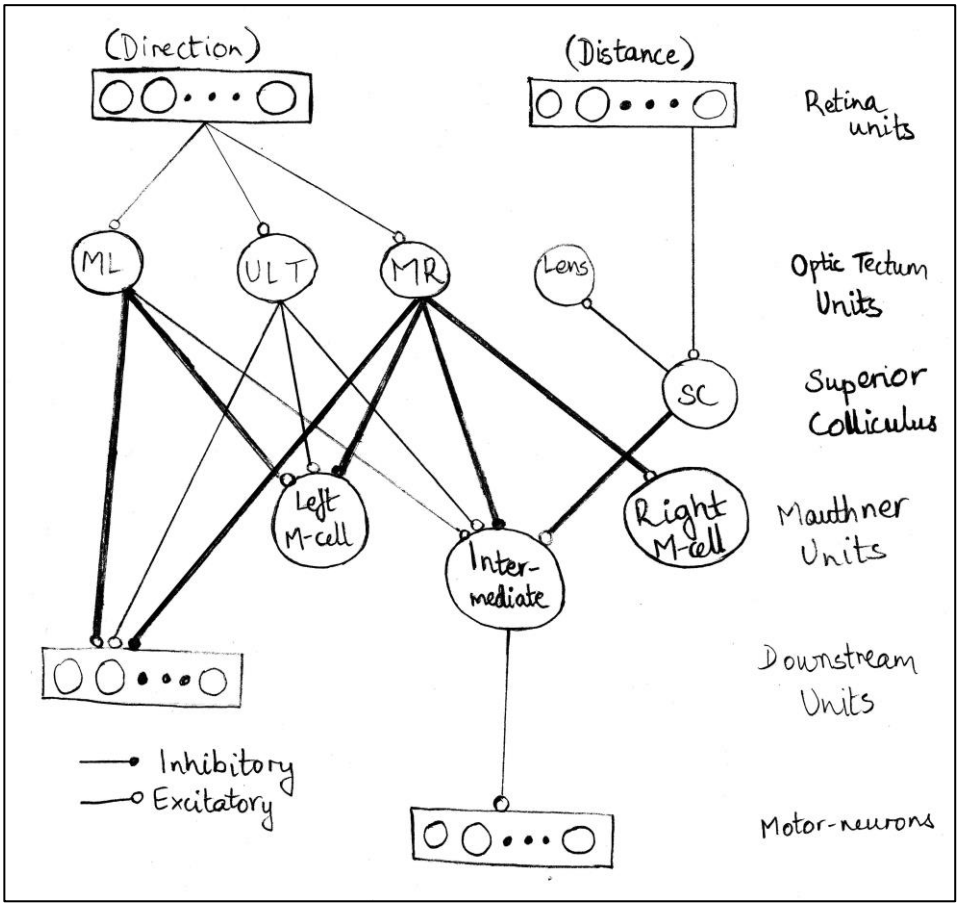


Figure 16: *Possible network that responds to the movement of prey so that fish can execute their predictive start. This network only accounts for the short-term response. Once the initial portion of the action has been carried out, higher command centres kick in to fine-tune the predictive start and ensure that the fish reaches the prey on time.*

The weights of the lines represent the strength of the excitation or the inhibition on the post-synaptic neuron. The network represents only the left side of the fish. Since the morphology is symmetric, the same network is present on the right side of the fish. In the diagram, there is an excitation shown from a left optic tectum unit to the Right M-cell. This has been included for completeness and the same applies from the right side to the Left M-cell (not shown). Some of the terms in the above network are explained below:

1. ML – movement of the prey from the left of the fish to its right
2. MR – movement of the prey from the right of the fish to its left
3. ULT – Ultra Low Threshold. This neuron fires to stimuli that have low amplitude
4. SC – Superior Colliculus
5. Intermediate – This stands for the intermediate neurons that receive information from the Superior Colliculus and the optic tectum units and integrate them before passing them on to the motor neurons
6. Lens – the optic tectum units that control the muscles that control the thickness of the lens in the fish's eye.

There are two pools of retina units, one that responds to the distance of the prey from the fish and one that responds to the horizontal trajectory of the prey. The direction pool has excitatory afferents on three optic tectum units, namely the MR, ML and ULT. The MR has strong inhibitory afferents on the Left M-cell, the Intermediate neurons and the Downstream units, and it has strong excitatory afferents on the Right M-cell. The ML has strong excitatory afferents on the Left M-cell and the Downstream neurons, and it has weak excitatory afferents on the Intermediate neurons. The Left M-cell computes the three converging inputs as follows:

1. If the prey is moving to the left, the Left M-cell should not fire, otherwise the fish would move in the opposite direction to the prey. Therefore, the MR strongly inhibits the Left M-cell, and without the excitation from the ML, the Left M-cell does not fire.
    a. There is a subsequent case should the prey be moving to the left. If the prey is already to the left of the fish and moving to the left, then it would be out of the visual field of the right eye. So, from the left optic tectum, the MR provides strong excitation to the Right M-cell and inhibits the Left M-cell, thereby making the fish turn left.
2. If the prey is moving to the right, then there would be strong excitation from the ML neuron and the Left M-cell would fire. There would be no inhibition from the MR neuron.
3. In the case that the prey is moving forward, there would be almost no excitation or inhibition from the MR and ML neurons. The ULT will be the only source of excitation and will make the Left M-cell fire weakly as the fish moves forward.
4. In the case that the prey moves caudally to the fish, the M-cell alone would be insufficient to turn the fish in the direction of the prey. Other downstream neurons would be required to make the fish turn more than it would with the M-cell alone. The faster or farther the prey is moving, the more signals will be sent along the ML or MR neuron. An important characteristic of the downstream neurons is that the further downstream they are, the higher their thresholds are, and increasing excitation is required for them to fire. So, more signals would need to be sent along the ML neuron to activate them. The frequency and number of action potentials being generated along the ML neuron will decide how many of the downstream neurons fire. I have stated that excitation from the ML is necessary on the left side of the fish because if the prey were moving to the left then the downstream neurons would be inhibited by the MR neuron.

This explains the response of the fish to the horizontal trajectory of the prey. The retina units responsible for determining the distance of the prey constantly keep feeding information to the Superior Colliculus. One of the ways that fish determine the distance of the object is the thickness of the lens in its eye. Therefore, I hypothesise that the same neurons that control the muscles controlling the lens also play a role in the predictive start of the fish. The Superior Colliculus receives information from the retina and sends information back to it via the deeper layers of the optic tectum. The thickness of the lens represents the distance of the object in question. To keep the lens thin to focus on objects far away, the muscles controlling the lens need to be stimulated to contract. I believe that the tetanus that is sent to the lens muscles compute distance of the prey and therefore can also be sent to an Intermediate neuron that will act as a logic gate, passing on the signals to motor neurons.

The computation that takes place in the Intermediate neuron is as follows:

1. The Intermediate neuron should fire only if the ipsilateral M-cell also fires. Therefore, the same inputs for the M-cell are also applicable to the Intermediate neuron.
2. However, since the information from the Superior Colliculus is more important than that received from the optic tectum, the afferent of the Superior Colliculus has a high weight attached to it.
3. If the prey is moving to the right, the MR will be inactive and the intermediate neuron will be activated by the ULT, ML and SC, but the firing of the Intermediate neuron will be controlled by the SC.
4. If the prey is moving to the left, the MR neuron will be active and inhibit the Intermediate neuron. Without the activity of the ML neuron, the SC alone cannot overcome the inhibition of the MR and so the Intermediate neuron does not fire.

It should be noted that the proposed network only accounts for the short-term processing. It does not account for the continuous processing that is required by a circuit that fine-tunes the movements of the fish in the predictive start. I believe that once the preliminary orientation of the fish has been carried out by this network, higher command centres are able to provide inputs to the network and further control the motion of the fish.

What is the benefit to humans from the extensive research being conducted on the neural networks of fish? As Stefan Schuster wrote in 2011:

> *"Complementing the flourishing study of decision making in the primate brain, fish can help us advance our understanding of sophisticated aspects of decision-making to the cellular and circuit level."*

For me, one of the most intriguing phenomena the work of these researchers has revealed is that the connections in our neural circuits inherently encode certain physical phenomena, like particle acceleration or angular image size. I think this could be extrapolated to human brains as well. The circuits in our minds might also inherently be encoding simple physical phenomena, which enable us to respond to certain stimuli. It is possible, therefore, that since these neural circuits are built to encode certain phenomena, they are identical across all humans.

The long-term potentiation studies of the fish also have vital implications for us. Most of us are completely oblivious to a lot of the stimuli around us. For example, I live quite close to an airport and airplanes fly past my house every three minutes. At first, I used to jump every time one flew past and rush to the window. However, within a couple of months, I was accustomed to them. Now, after twelve years, I barely notice them anymore. Just like the fish, long-term potentiation may have been at work in my inhibitory pathways and that has now numbed me to the constant stimulus of the airplanes. There are millions around the globe who face similar situations, where they are indifferent to some constant stimuli in their environment. The studies conducted into the decision making process in teleostei may have provided an answer as to how this habituation is represented in our neural circuits.

Although some argue that knowledge is sufficient in and of itself, I believe that it is not good unless we do something with it. We can use the data obtained from the study of neural circuits in fish to further the development of Artificial Intelligence (AI). As of now, most AI is based on deductive reasoning and is limited by the protocols that the programmer has coded into the system. If circuits that mimic the neural networks of the lower order vertebrates are developed, it would be quite a significant leap forward in AI. If conscious efforts were made to inculcate some of the predominant structures in our neural networks into the circuits of machines, their computing power might increase drastically. Machines

might not just be restricted to what they have been programmed to do but might also learn on their own based on repeated instructions presented to them.

The neural networks in the brains of humans and other higher order vertebrates are currently too difficult to access, or there may be some ethical issues with regard to implanting electrodes in live humans and then subjecting them to stress tests. Therefore, researchers have chosen to explore the neural networks in teleostei, specifically zebrafish. It is hoped that these studies will reveal some important mechanisms of function that can then be adapted to better understand human neural networks. Rather than waste time and money blindly testing human beings for a variety of responses, perhaps these studies on zebrafish could guide researchers to the most productive experiment to conduct on human beings.

References

B., P. (2013, June 5). I-Search: Visual Aid: The Nervous System. Retrieved July 14, 2016, from https://prezi.com/5je2eahdig1i/i-search-visual-aid-the-nervous-system/

Bellman, K. L., & Krasne, F. B. (1983, August 19). Adaptive Complexity of Interactions Between Feeding and Escape in Crayfish. *Science, 221*(4612), 779-781. doi:10.1126/science.221.4612.779

Bene, F. D., Wyart, C., Robles, E., Tran, A., Looger, L., Scott, E. K., . . . Baier, H. (2010, October 28). Filtering of Visual Information in the Tectum by an Identified Neural Circuit. *Science, 330*(6004), 1147-1160. doi:10.1126/science.1192949

B., P. (2013, June 5). I-Search: Visual Aid: The Nervous System. Retrieved July 14, 2016, from https://prezi.com/5je2eahdig1i/i-search-visual-aid-the-nervous-system/

Canfield, J. G., & Rose, G. J. (1993, May). Activation of Mauthner neurons during prey capture. *J Comp Physiol A Journal of Comparative Physiology A, 172*(5), 611-618. doi:10.1007/bf00213683

Canfield, J. G., & Rose, G. J. (1996). Hierarchical Sensory Guidance of Mauthner-Mediated Escape Responses in Goldfish (*Carassius auratus*) and Cichlids (*Haplochromis burtoni*) (Part 2 of 2). *Brain Behav Evol Brain, Behavior and Evolution, 48*(3), 137-156. doi:10.1159/000316283

Eaton, R. C., Canfield, J. G., & Guzik, A. L. (1995). Left-Right Discrimination of Sound Onset by the Mauthner System. *Brain Behav Evol Brain, Behavior and Evolution, 46*(3), 165-179. doi:10.1159/000113269

Faber, D.S., and Korn, H. (1978). Electrophysiology of the Mauthner cell: basic properties, synaptic mechanisms, and associated net- works. In Neurobiology of the Mauthner Cell, D.S. Faber and H. Korn, eds. (New York: Raven Press), pp. 47–131.

Gahtan, E., & O'Malley, D. M. (2003, April 10). Visually guided injection of identified reticulospinal neurons in zebrafish: A survey of spinal arborization patterns. *The Journal of Comparative Neurology J. Comp. Neurol., 459*(2), 186-200. doi:10.1002/cne.10621

Guzik, A. L., Eaton, R. C., & Mathis, D. W. (1999, March). A Connectionist Model of Left-Right Sound Discrimination by the Mauthner System (G. Laurent, Ed.). *Journal of Computational Neuroscience, 6*(2), 121-144. doi:10.1023/A:1008828501676

Korn, H., Oda, Y., & Faber, D. S. (1992, January 01). Long-term potentiation of inhibitory circuits and synapses in the central nervous system. *Proceedings of the National Academy of Sciences, 89*(1), 440-443. doi:10.1073/pnas.89.1.440

Korn, H., & Faber, D. S. (2005). The Mauthner Cell Half a Century Later: A Neurobiological Model for Decision-Making? *Neuron,47*(1), 13-28. doi:10.1016/j.neuron.2005.05.019

Liu, K. S., & Fetcho, J. R. (1999). Laser Ablations Reveal Functional Relationships of Segmental Hindbrain Neurons in Zebrafish.*Neuron, 23*(2), 325-335. doi:10.1016/s0896-6273(00)80783-7

M. (n.d.). The generation of action potential in nerves. Retrieved July 14, 2016, from http://www.animalresearch.info/en/medical-advances/nobel-prizes/the-generation-of-action-potential-nerves/

O'malley, D. M., Kao, Y., & Fetcho, J. R. (1996, December). Imaging the Functional Organization of Zebrafish Hindbrain Segments during Escape Behaviors. *Neuron, 17*(6), 1145-1155. doi:10.1016/s0896-6273(00)80246-9

Oda, Y., Charpier, S., Murayama, Y., Suma, C., & Korn, H. (1995, September). Long-term potentiation of glycinergic inhibitory synaptic transmission. [Abstract]. *J.*

Neurophysiology, 74(3), 1056-1074. Retrieved June 13, 2016, from http://www.ncbi.nlm.nih.gov/pubmed/7500132

Oda, Y., Kawasaki, K., Morita, M., Korn, H., & Matsui, H. (1998, July 09). Inhibitory long-term potentiation underlies auditory conditioning of goldfish escape behaviour. *Nature, 394*(6689), 182-185. doi:10.1038/28172

Preuss, T., Osei-Bonsu, P. E., Weiss, S. A., Wang C., & Faber, D. S. (2006, March 29). Neural Representation of Object Approach in a Decision-Making Motor Circuit. *Journal of Neuroscience, 26*(13), 3454-3464. doi:10.1523/jneurosci.5259-05.2006

Schlegel, T., & Schuster, S. (2008, January 04). Small Circuits for Large Tasks: High-Speed Decision-Making in Archerfish. *Science,319*(5859), 104-106. doi:10.1126/science.1149265

Schuijf, A. (1975). Directional hearing of cod (Gadus morhua) under approximate free field conditions. *Journal of Comparative Physiology? A J. Comp. Physiol., 98*(4), 307-332. doi:10.1007/bf00709803

Schuster, S. (2012, April). Fast-starts in hunting fish: Decision-making in small networks of identified neurons. *Current Opinion in Neurobiology, 22*(2), 279-284. doi:10.1016/j.conb.2011.12.004

Swanson, B. O., & Gibb, A. C. (2004, November 01). Kinematics of aquatic and terrestrial escape responses in mudskippers. *Journal of Experimental Biology, 207*(23), 4037-4044. doi:10.1242/jeb.01237

Weiss, S. A., Preuss, T., & Faber, D. S. (2009). Phase Encoding in the Mauthner System: Implications in Left-Right Sound Source Discrimination. Journal of Neuroscience, 29(11), 3431-3441. doi:10.1523/jneurosci.3383-08.2009

Wohl, S., & Schuster, S. (2007, January 15). The predictive start of hunting archer fish: A flexible and precise motor pattern performed with the kinematics of an escape C-start. *Journal of Experimental Biology, 210*(2), 311-324. doi:10.1242/jeb.02646

Zottoli, S. J., Hordes, A. R., & Faber, D. S. (1987). Localization of optic tectal input to the ventral dendrite of the goldfish Mauthner cell. *Brain Research, 401*(1), 113-121. doi:10.1016/0006-8993(87)91170-x

Glossary

Axon – the projection of the neuron that carries signals away from the cell body and to neurons or body cells.

Azimuth – the direction of an object in space expressed as angular distance from a predetermined fixed point. Usually used in the field of astrology.

Caudal – towards the tail of the organism

Contralateral – on the opposite side of the organism.

C-start escape – reflex action of teleostei in response to stimuli. Body turns to resemble a 'C.'

Dendrite – the projection of the neuron that receives signals from other neurons or from sensory cells.

Depolarization – the trans-membrane potential becomes less negative and may result in an action potential.

EHP – Extrinsic Hyperpolarizing Potential. It is the field generated by an IPSP at the M-cell axon cap.

EMG – Electromyography. It is a form of diagnosis used to assess muscle and motor neuron health.

Endorgan – an encapsulated ending found in sensory neurons.

EPSP – Excitatory Postsynaptic Potential.
Hyperpolarization – the trans-membrane potential becomes more negative.
IPSC – Inhibitory post-synaptic current.
Ipsilateral – on the same side of the organism
IPSP – Inhibitory Postsynaptic Potential.
Latency – the delay time between the stimulus onset and the response
Lesion – damage to an organ or tissue
M-cell – Another name for Mauthner cell
M-Spike – the action potential generated by the M-cell.
Neurotransmitter – chemical released into a synapse between two nerves.
Optic tectum – The neuron that transmits information from the retina to the M-cell system and the Superior Colliculus
PHP neuron – neurons that have inhibitory synapses on the M-cell.
Resting Potential – the baseline potential of the neuron when it is not firing
Rostral – towards the head of the organism
Saccular Hair Cell – hair cell in the inner ear of the fish responsible for reacting to pressure.
Soma – the main body of the cell
Suprathreshold – Exceeding the threshold of the neuron
Synapse – the part between two nerve cells, where the axon of one nerve ends on the dendrite of another.
Teleost – in classification, it is the class that contains ray-finned fish.
Threshold potential – the potential at which the cell membrane must depolarise in order for it to fire.
Transmembrane – across the membrane of the cell
Utricle Hair Cell – hair cell in the inner ear of the fish responsible for reacting to particle motion.

From the Dark Knight to Francis Underwood: Twenty-first Century Noir Heroes

Lu Zeng

Author background: Lu Zeng grew up in China and currently attends Shenzhen Foreign Languages School, located in Shenzhen, China. Her Pioneer seminar topic was in the field of film studies and titled "Film Noir and its Contexts."

I. Introduction

A decade after the Cold War ended, the United States had not yet found a new aggressor to resist, and the motion picture industry had not yet found a new direction to follow. Just as the U.S. intelligence agencies were groping for a new rival, two planes from Al Qaeda hit the World Trade Center in New York City. The whole world was shaken. And there was the new target: terrorists.

Entering the 21st century, world citizens are increasingly concerned with the rise of terrorism and possible failures of existing social systems. Over the last decade, 9/11 and other fearmongering terrorist attacks deeply inserted cultural trauma into people's consciousness. Media and culture not only respond to the cultural trauma we have today but also alter the way we perceive it. Ann Kaplan has extensively studied the politics and psychology of cultural trauma. In her introduction to *Trauma Culture*, she points out the significance of media's effect on trauma: "People encounter trauma by being a bystander, by living near to where a catastrophe happened, or by hearing about a crisis from a friend. But most people encounter trauma through the media, which is why focusing on so-called mediatized trauma is important." (Kaplan 2)

Subjects of classical Hollywood have begun to die away: instead of romantic comedies and inconsequential drama movies, more silver screens around the globe are now displaying political players overriding laws, superheroes saving the world from nuclear destruction, and crime masters taking advantage of every loophole of established systems. Yet twenty-first century audiences have had no trouble favoring characters with flawed personalities or those who make sordid attempts at gaining personal interests; the popularity of such screen characters unsurprisingly remind me of the pessimistic, lonely noir heroes from film noirs of the 1940s and 1950s. Classical film noir was the "dark cinema," those that have unsentimental male leads, chiaroscuro lighting effects, and femme fatales who jeopardize the detective's lead character's unlawful plans.

Of all the noirish twenty-first century characters that have gained success either on the silver screen or television, the Dark Knight created by Christopher Nolan and Francis Underwood from the Netflix series *House of Cards* are the two that best construct the concept of twenty-first-century noir heroes.

II. Noir and Terror in the Twenty-First Century

"We don't submit to terror. We make the terror."[1]

—*House of Cards* Season 4

In the popular Netflix production *House of Cards,* the President of the United States, Francis Underwood (played by Kevin Spacey), alongside his unsmiling wife Claire (Robin

Wright), decides to declare war on ICO (Islamic Caliphate Organization), an Islamic terrorist organization which is unmistakably a fictional equivalent of ISIS (the Islamic State of Iraq and Syria). However, this decision is by no means a determination to fight terrorism or to secure American citizens' safety. Francis Underwood paints ICO as a threat that the United States has to counter, while it is a gesture made completely out of insecurity. Underwood, in fear of the soaring approval ratingof his competitor in the ongoing Presidential election, is determined to use his strong, intimidating role in the fight against ICO to turn the tables around to secure the Presidency he initially obtained by forcing the then president to resign from office.

Millions of lives are substantially threatened by rising terrorism; however, in this narrative, the head of the state which honors freedom and democracy chooses to sacrifice an unnecessary number of soldiers to terrify his own citizens. The terror such groups as Al Qaeda and the Islamic State produce makes us wonder where the next attack is going to be, and who will be there to protect us? Depressingly, terrorism is only spreading faster and more broadly, with Al Qaeda accusing ISIS of being overly brutal. Efforts to combat terrorism have not achieved substantial results; in fact, governments seem weak before terrorist organizations. European capitals like Paris and Brussels have been ravaged in the past year; masses of land in the Middle East has been taken by ISIS; more and more Western youths are joining ISIS each year.

As demonstrated by television shows like *House of Cards*, the ineffectuality that governments show in eliminating terrorist threats triggers increasing public distrust for authorities. The expanding influence of terrorism brought with it the questioning of existing social systems. The powerlessness people feel in the face of terrorist aggression is inversely reflected by the power of characters on screen, for instance, superheroes – fantasy figures that lessen the sense of vulnerability.

Superhero movies, well represented by Christopher Nolan's Batman reboots, appeal to the audience by providing heroic characters who are individually capable of saving the world from being annihilated by terrorists. In the meantime, hit television series such as *House of Cards*, *The Blacklist*, and *Person of Interest* depict either the decay of bureaucracy or unrealistic solutions to improve present government agencies. In the twenty-first century, noir is an element pervasive in television and film that continues to communicate "social problems" and serve as a novel ingredient for providing escape. Superheroes assure their audience that their homes will not be destroyed by terrorists, while others push the audience to face their concerns about ruling governments. Film noir, having evolved over the years, has replaced the cheerful, optimistic style that once dominated Hollywood and television. In the form of neo-noir, it is now everywhere. As Wheeler W. Dixon says about the noir in twenty-first century entertainment, "This is the true 'reality' of life in the twenty-first century; rapacious greed, fear, violence, endless war, terrorism, and the continual droning of either threats or assurances from impotent authority figures, interested only in their own acquisition of power." (Dixon 2005:153)

The mounting fear produced by terrorism has inevitably led the masses to question the authorities. Are they capable of protecting a nation? Are they even capable of protecting a city? The Dark Knight responds to the rise of terrorism, while Francis Underwood gives his own answer to the last two questions. A remarkable similarity shared by the two of them is the noir elements attached to their characterizations, which is worth a comparison with noir characters from the original noir cycle.

III. The Dark Knight: Noir Hero Rising with Terrorism

> After 9/11, stories about the end of the world were rewritten for kids, while stories about saving it turned dark, for adults […] Superheroes present a reversal of war-on-terror narratives that reimagine monsters as alluring—here, heroes are outsiders, even if they have audience allegiance and sympathy. And if young adult dystopia is in the end ironically optimistic, are the never-ending sequels and never-ending fights in superhero films signs of hope in the war on terror, or futility? (Kavadlo 158)

In the past few years, superheroes on screen who saved Marvel and DC Comics from bankruptcy also saved the world over and over again. Comic-book movies are now a huge business: Iron Man, Batman, the X-Men, Spiderman, and other comic-book heroes and heroines—the darker the better (Dixon 2011:43), for the present moment. Before the strike of terrorism caused pervasive cultural trauma, heroes also existed for collapsing worlds, but the danger portrayed was never so dismal. In today's superhero movies for the first time, capitals of human civilization are destructible. Although the audience knows well that the world will be saved in the end, things can barely return to their original state before damage. When heroes are effortlessly fighting the "bad guys," the collateral damage is usually colossal: cars are crashed, buildings are wrecked, streets are ravaged, and everything is detonated. The level of darkness of these films is continually escalating, and advancing cinematic technologies have enabled them to be perfectly convincing to the audience. The audience seeks excitement in the confrontation between the good and the evil, and the darker the movie, the more excitement.

Christopher Nolan's Dark Knight movies are by far the darkest of the genre. Recalling classical film noirs, the trilogy has a noirish visual style. In the second half of *Batman Begins*, Gotham City, the movie's symbolic New York, is depicted as a dark, rotten city – an urban wasteland as in classical film noirs. The streets are narrow, damp and filthy, and the scenes are predominantly set in the night. In such a noir setting, Bruce Wayne, the Dark Knight, puts on his black Batman suit and sets out into the night. The whole city is in terror because of the fear of toxin placed by Ra's Al Ghul, Wayne's spiritual mentor who was previously disguised as his own assistant, and a character directly based on the world's most wanted terrorist: Osman Bin Laden.[2] His motive for destroying Gotham is precisely that of any extreme terrorist. He sees himself as the one to "restore the balance" of Gotham because it has reached the "pinnacle of decadence," and he does not have a doubt about letting Gotham "tear itself apart in fear." (Nolan 2005)

In the second of the trilogy, *The Dark Knight*, elements of noir go beyond the visual level. Everything is chaotic. This time, the "bad guy" is the Joker, a character whose appearance is already terrifying, despite its clown-like aspect. Nobody thinks he would be a real threat when he walks into the room of his wealthy, powerful employers: shaggy hair, cheap purple suit, and a comic joker face. He is another terrorist figure, killing in the fashion of a serial killer, and is adept at creating mass terror by causing damage arbitrarily. He calls himself, an "agent of chaos." Apart from challenging Batman, he does not have any clear, justified motive, and seems to just "want to watch the world burn." Criminals in Gotham have been intimidated by Batman and the streets of Gotham are cleaned up. However, the Joker is a stronger opponent than Ra's Al Ghul. Ra's Al Ghul had an ideal he was devoted to, but the Joker is not a man of principles. He has a world of rules to break and nothing to believe in, which makes him much harder to be defeated. Although Batman is a heroic figure and the protagonist in this film, he loses his superpower appeal in the confrontation

with the Joker. The same happens in *The Dark Knight Returns*, in which a nuclear bomb almost kills the Dark Knight.

At the very beginning of *Batman Begins*, Nolan introduces the fear and the trauma that are the basis of the characterization of Bruce Wayne. Fear weakens him, while trauma strengthens him. Both originate from Bruce's childhood. *Batman Begins* opens with young Bruce falling into a deep well when playing with his friend Rachel. He was alone, frightened and hopeless. The group of bats flying past his face only intensifies his emotions; they are scary and seemingly invincible. Though he was saved by his loving father, bats remain an unconquerable fear of his until he becomes Batman. Viewers know that Bruce will eventually appropriate the image as his own, to terrify those who would terrify others. But in the beginning, bats symbolize Bruce's pain and powerlessness, abandonment later on, and fear itself. (Kavadlo 164)

His trauma also comes from his fear of bats. When young Bruce was taken by his parents to a show, he was disgusted by the presentation of bats. He had to ask his parents to exit early, but just a few moments after they left the theater, his parents were robbed and killed by a gun. He blames himself for his parents' death, but more of the time he is angry. In the hope of getting revenge for his parents, he travels the world to study the minds of criminals, trying to understand his enemies. But he has become truly lost. He then meets Henri Ducard in a prison, the "assistant" of Ra's Al Ghul, "a man greatly feared by the criminal underworld" and the head of the League of Shadows. He offers him a path to serve "true justice," and Bruce takes it, expecting to learn the tactics needed to save his city from decadence. He was trained and approved to leave, making his grand return to Gotham as the heir to Wayne Enterprise, his family business. But that is his identity by day; by night, he masks himself to become Batman, a vigilante who is not "just a man lost in the scramble for his own gratification," as Ducard describes vigilantes when they first meet.

When Bruce Wayne graduates from the League of Shadows and comes back to his rotten city to save it, he is already some kind of a noir hero. A loner with a "dark" past, as Naremore puts it. (Naremore 250) Loneliness and darkness are two key features of Bruce Wayne that help us read the Dark Knight, who eagerly wants to protect his city from corruption and crime.

As a vigilante working without the authorities, he acts alone, refusing any real partners, just like the private eye in classical noir. In the opening scene of *The Dark Knight*, Batman ties up three imposters of him with the criminals and replies to their offer to help with four firm, simple words: "I don't need help." He works closely with the Police Department, but always at a distance. Lieutenant Gordon, and later Harvey Dent, the District Attorney of Gotham City, find him highly unpredictable. They can never contact him directly; Batman appears and disappears anytime he feels the need to. Batman shares plans with the authorities, but more often he keeps secret plans to himself. Even Lucius Fox, the head of the Applied Sciences Department of Wayne Enterprise and Bruce's confidant, does not gain his full trust. Bruce Wayne reassigns the Applied Sciences Department to a government telecommunications project to start a secret project to help him track criminals, but he does not notify Fox. When Fox asks about the government contract, Bruce light-heartedly says, "I'm playing this one pretty close to the chest." In fact, the only one who is close enough to his chest is he himself.

He has been lonely all along since he decided to pursue justice, initially for his parents but later for Gotham. Before he realizes that revenge is not going to make any real difference, he was determined to kill Joe Chill, his parents' murderer. But after Chill is killed by a mob leader before him, Bruce shifts from personal vengeance to public protection. (Kavadlo 165) Enlightened by Ra's Al Ghul, he comes tounderstand that it is the

troubled city that killed his parents, not Joe Chill alone. On the risky path to serve true justice, he tries to implicate as few people as possible.

For the safety of others, he must be masked. He must become someone else. The real identity of Batman can never be revealed. Batman might be omnipotent, but Bruce Wayne is only a young billionaire who has so much to lose. However, in the face of a criminal like the Joker, Batman is easily revealed, and so is Batman's weakness: Rachel. Bruce has loved her since childhood, but she could not stand the double life Bruce is living. She promises Bruce that they can be together when Gotham does not need Batman, but just prior to her death, she chooses to be with Harvey Dent, the white knight of Gotham. Just when Bruce makes up his mind to put down the Joker after a brief period of wavering, Rachel and Harvey Dent are abducted as a part of the Joker's plan to break down Batman and Gotham. Sadly, in the end, Bruce saves the wrong person. Rachel, Bruce Wayne's one and only love, is killed in an explosion. Similar to classical noir characters, Batman can never settle down or find love. (Kavadlo 165)

As his name makes clear, the Dark Knight is dark. Despite being a heroic figure, he is morally ambiguous. In *Batman Begins*, Ra's Al Ghul directs Bruce Wayne's path to "serve true justice": "To conquer fear, you must become fear . . .you must bask in the fear of other men...and men fear most what they cannot see. It is not enough to be a man . . . you have to become an idea . . . a terrible thought . . . a wraith." And Bruce Wayne takes his advice, embraces his worst fear, and becomes one with the darkness. Pushed by Ra's Al Ghul and Alfred, Bruce Wayne becomes Batman, embracing the fear that he has harbored for most of his life: bats. But however he believes in his cause, he is a vigilante, only one that has a greater purpose. A vigilante is by no means the embodiment of American ideals. Rather, his endeavors are a violation of the judicial system. He individually counters terrorists, while utilizing terror as his weapon. When he is put in a room with the Joker to interrogate him, he uses extreme, even unlawful violence to make the Joker talk. Police officers who are watching them could barely sit by. He is not stopped only because James Gordon, the lieutenant who personally trusts Batman, silently approving his use of violence.

"You complete me…You are just a freak like me." (Nolan 2008) In this most arresting scene of all interactions between Batman and the Joker, the Joker gives his morbid but heartfelt confession to Batman. Indeed, the Joker is not just a foil to Batman. Batman and the Joker are both people who want to single-handedly change the world, only working in opposite directions. The Dark Knight employs terror on criminals, but the extent to which he uses it inevitably produces an antagonist like the Joker.

The Joker is a sociopath who claims to be "ahead of the curve," mocking the morals and codes of the authorities. But he could not be simply classified as "the evil." The moral ambiguity exists both in the Joker and the Dark Knight. The Joker somehow wants to reveal to the public how useless moral codes and rules are by cracking the minds and belief of people. In a sense, he believes in a different but not entirely incomprehensible justice. "The only sensible way to live is without rules," he concludes to the Dark Knight. To convince the world, he craves power and attention. Every damage he causes is to create social disorder, not to eliminate a city or civilization, like Ra's Al Ghul does.

The Dark Knight, by contrast, is willing to pay any price to maintain social order. He does not even want to be a hero. He can sacrifice himself and be the villain, if it can help Gotham get rid of crimes or threats. However, even though he appears to be a selfless, righteous knight of Gotham, he is an outlaw who takes advantage of people's fears. The morally ambiguous nature of both of them makes it reasonable to think of them as the "angelic killers" James Naremore presents in his discussion of the film noirs adapted from Graham Greene's novels. Like Harry Lime (played by Orson Welles) in *The Third Man*, the

Joker and the Dark Knight commit crimes, yet their sanity and self-justification render moral ambiguity in them.

IV. Francis Underwood: Impeacher of the Republic

> With the decline in cinema attendance in the early 2000s eerily mimicking the same pattern in the early 1950s (with the dawn of television, and the advent of the Cold War), television programming has become a new and potent source of noir. *CSI* unravels grisly mysteries on a weekly basis; *Law and Order: Special Victims Unit* deals in brutal sex crimes and murders with studied detachment, even nonchalance; Fox's utterly meretricious *Moment of Truth* hooks its contestants up to a lie detector, and then asks them increasingly invasive questions about their personal lives, for ever larger sums of cash. (Dixon 2005:153)

Television in the twenty-first century has become unusually realistic. Recalling the definition of "reality" in the twenty-first century given by Dixon, "realistic" can now be considered a synonym of "dark." Not only are there reality shows like *The Apprentice* and *Survivor*, but regular series are also turning dark. *House of Cards*, the 2013 Netflix series has gained popularity around the globe. With a budget of 4.5 million dollars per episode,[3] it is the second most expensive series of Netflix. The sheer size of the budget only helps it become an even more "realistic" representation of life at the White House. Unlike the CBS series *Madam Secretary*, it is a dark and sometimes hopeless political drama.

Setting aside the actual contents of the series, *House of Cards*, like many social problem pictures in the past, has a noir-like visual style. Even without the iconic night scenes of classical noir, it tries to acquire a black-and-white look. The saturation of frames is always lower than normal, creating a subdued atmosphere throughout the show. Especially in the White House, the lighting is seldom sufficient, as if suggesting the White House is itself a dark, murky place. The darkness is intensified in Francis' vivid hallucinations when he is waiting for a liver transplant. Zoe Barnes, the journalist he killed, seduces him on the couch in the Oval Office, and there, too, Peter Russo shoves his face up against the window. Later in the season, Nancy Durant, Francis' Secretary of State, is completely intimidated by his description of this disturbing scene: "Cigarettes and razor blades and sex. It was terrifying." The highlight in those frames is all tinted blue, and the rest is bleak shadow. The underexposure also adds to the effect; the hallucinations seem even more dreary.

Like the classic White House drama *The West Wing*, the realism of *House of Cards* stems from its timeliness. Lynn Spigel, Professor of Screen Cultures at the School of Communication at Northwestern University, comments on *The West Wing*: "… perhaps more than any other network series, it derives its 'quality' appeal from its 'timely relevance' and deep, if melodramatic, realism." (Spigel 242) The most recent example would be Season 4, released in March 2016, which has an apparent base in the ongoing presidential election in the United States. Although the presidential election in the show does not involve direct parallels to the candidates in the actual 2016 US election, the means that candidates employ to win votes are similar, for example, utilizing social media. Will Conway, Francis Underwood's top competitor, is a king on social media. He draws public attention by exposing his own private life, posting pictures and videos of his family and his daily activities. The only difference between Conway and Donald Trump on social media is that Conway is building an ordinary family-man image. However, his scheme goes beyond merely earning people's affection; he wants to point directly to people's concerns. To achieve that, he illegally works with a search engine company, collecting and analyzing

people's search questions. Soon after Underwood becomes aware of Conway's secret, Frank makes sure it is revealed. This can be interpreted as suggesting the same phenomenon taking place in the actual presidential election, which would be a darker truth.

Francis Underwood, the protagonist of *House of Cards*, is a downright egotistic, duplicitous politician. From Season 1 to Season 4, he climbs his way up from the United States Congress to the Oval Office, without conforming to usual processes of democratic elections. His rise to power is indeed corrupt. He kills Peter Russo and Zoe Barnes in order to secure his position of Vice President, and plots against the President of the United States to force him to resign under the threat of impeachment. He thereby rises to President automatically, being the first on the list of succession. He does not hesitate to override the law when it can serve his self-interest. He has always been a cold, emotionless character, not only as a politician, but also as a husband.

His relationship with his wife is an utterly peculiar one. Claire Underwood has been a cold, shrewd conspirator of her husband's for thirty years. As the New York Post author Robert Rorke describes, "Claire is one of the nastiest characters to walk across a TV screen. She speaks only in veiled threats (or outright) threats, keeping her voice at a temperature somewhere between freezer-burn-level and a slow ice-cube-thaw."[4]As mentioned earlier, *House of Cards* has a marked low visual saturation. The low saturation of the frames fits the character of Claire Underwood especially well. Her clothing is equally monochromatic, and her face is always paled by the visual effect, thus communicating the coldness under her skin. When they get married, they are both very clear what they are going to achieve with their marriage. The conventional family life with children and a big house means nothing more than mediocrity to them. They want power. They are more than just life partners; they are also career partners, "learning from and advising each other," as Claire puts it. From Season 1 through Season 3, Claire Underwood gives up her previous career in order to help Francis rise to the Oval Office.

Their marriage is also never bound by the rule of loyalty. They both have had affairs, and they are both aware of the other's affairs, but neither of them ever sees the affairs as any obstacle to their marriage. Their common goal transcends the simple meaning of being married. As much as Francis loves Claire, he understands he cannot give her what a lover would. Probably the best example would be Claire's affair with Thomas Yates, the couple's biography writer. In the estrangement between Claire and Francis, Yates won Claire's affection by his brilliance and gentleness. Francis soon becomes aware, but instead of opposing the affair, he blesses Claire and Yates because "one person cannot give everything to another person," as if providing a partner for what she wants in order to keep her content with cooperating. The *Variety* review accurately comments on Francis's choice: "Frank's blessing of Claire's affair made clear, their partnership has morphed into an entirely professional exercise, with a shared lust for power having supplanted more conventional matrimonial bonds."[5]

Even with such a fierce character, like the classical noir hero, Francis Underwood can feel helpless. After he reaches the Oval Office, which can be seen as the ultimate goal of all previous shady efforts, the real challenges begin. Underwood who has never feared anything or anyone is now for the first time afraid: afraid of losing his job and afraid of losing his wife. Tom Hammerschmidt, former editor-in-chief of *The Washington Herald*, is, in the new season, dedicated to exposing Frank's path to presidency, which contains enough dirt to sabotage the entire political career of Frank. Even his wife cannot sit tight. Claire craves power, but Francis has trouble granting it to her. Claire has become a danger to him; they have not slept in one bed for a long time, and he continuously imagines himself hitting the image of Claire in the mirror. Their conflicts lead to a long, destructive estrangement. Not

only does Claire have an affair with Thomas Yates, she tries to ruin her husband's election by putting up Frank's father's picture with the Ku Klux Klan on billboards. Finally, however, she comes back to team Underwood after a game changing event – the attempted assassination of Francis.

Lucas Goodwin, boyfriend of the dead Zoe Barnes, wants Francis Underwood dead more than anything. The unsuccessful assassination put Francis in a coma and kills Meechum, the Underwoods' beloved secret service agent. Claire then finally returns to her husband to help him get through the crisis while he is still in a coma. In order to keep her on board, Francis decides to let Claire run for vice president as his running mate. Though she is back for cooperation, the relationship between her and Francis becomes even icier. She is fully armed, ready to go to war with her husband at any time. Love is not what glues them together; more likely it is the greed for power. After she returns to the team, she studies the way to beat the Conways late into the night. She finds a video recording their first encounter with the Conways years ago, and tells Francis about it over the phone. After watching the video, Francis comes to her bedroom, and the only thing he says is, "We're going to destroy them." Claire replies, "Yes, we are." The greyish look of the bedroom, the blue-tinted highlight, and the Underwoods' emotionless faces make them seem like two engaged gears in a gargantuan mechanism.

Their marriage is very much like the relationship between Walter Neff and Phyllis Dietrichson in *Double Indemnity*. They are first brought together by love, but the common goal is what keeps them together. In *Double Indemnity*, it is killing Tom Powers; in *House of Cards*, it is acquiring power. But in the end, Neff and Dietrichson try to kill each other, both seeing each other as a threat. The undercurrent of grappling with each other also exists in the Underwoods' relationship all along. There are numerous times Claire threatens her husband, and the reverse is true. In this sense, Claire Underwood can be seen as a femme fatale herself, just like Phyllis Dietrichson, sometimes seductive but always independent, goal-oriented and strong.

House of Cards is an elaborate response to rising terrorism and destabilizing international society. If the Dark Knight is saving the collapsing world, then Francis Underwood is extracting the most profit while making the world worse. The Dark Knight tries to help the audience overcome its fears, while Francis Underwood lets the audience face the brutal truth of their fears. The success of Nolan's Dark Knight trilogy is rooted in the common desire to escape real life, but "there remains a substantial audience for cynical, sometimes darkly humorous entertainment leavened with a kind of glamour." (Naremore 298) The impact of film noir in this type of entertainment is undeniable. The harsh realism and dark relationships in classical noir films continue to exist today, serving to deliver dire depression.

V. Conclusion

In the twenty-first century, media functions as a channel of cultural trauma. Two diverging paths emerge in media's response to the trauma caused by terrorist attacks and destabilized domestic societies. One path is to reassure the viewers with superheroes on the screen and offset the powerlessness they experience when faced with nonstop attacks all around the world. The other path does the opposite: by pointing to the very fears of the public, it amplifies the cultural trauma in order to strike viewers with extreme realism like classical noir did in the twentieth century. In both paths, there are noir elements and characters with resemblance to classical noir heroes.

Nolan's Dark Knight series is a representative of the stream of superhero movies in the twenty-first century that follows the first path. The Dark Knight is noirish in multiple

ways: he is lonely, dark and sometimes powerless, though ultimately a hero that saves the world. On the contrary, characters like Francis Underwood in *House of Cards* show the darkest possibility of truth. Francis Underwood is an even closer analogy to the noir hero. He is cold, duplicitous and by all means dark.

The Dark Knight and Francis Underwood are both twenty-first century noir heroes, one saving the world and the other destroying the world. The Dark Knight movies help the audience get rid of the wounding fear of terrorists, whereas *House of Cards* pushes the audience to the truth that heads of states are corrupt and that they cannot provide security for them. Together they construct the present-day universe of trauma culture. One tries to heal the trauma, the other tries to worsen the trauma. They are two opposite views of the trauma we harbor today, and both have gained considerable popularity. They are the new faces of the classical noir hero, and the new generation of noir heroes will likely lead to an evolution of the genre of film noir in the new century.

Right now in the year of 2016, a messy election in the United States challenges people's belief in justice and freedom, conflicts of interests arise in the Syrian civil war, and terrorist attacks continue to prevail all around the globe. In this case, noir may become a more universal element in not only cinema, but also other forms of media. Cultural trauma is going to get closer to media spectators through noir, however, the fact that media can influence the weightiness of trauma shall not stand in the way of creating noir heroes with full imagination.

Notes

1. This line is excerpted from episode 13 of *House of Cards* Season 4.
2. See Kavadlo 166-167.
3. Please refer to the data listed at http://www.cheatsheet.com/entertainment/the-10-most-expensive-tv-shows-of-2015.html/?a=viewall.
4. See the article by Robert Roke at http://nypost.com/2016/03/17/the-most-shocking-moments-from-house-of-cards-season-4-2/.
5. See the article by Brian Lowry at http://variety.com/2016/tv/columns/house-of-cards-review-kevin-spacey-robin-wright-complete-fourth-season-binge-viewers-spoilers-1201725921/.

References

Dixon, Wheeler Winston. *21st-Century Hollywood Movies in the Era of Transformation*. N.p.: Rutgers UP, 2011. Print.

Dixon, Wheeler W. *Film Noir and the Cinema of Paranoia*. N.p.: Edinburgh UP, 2009. Print.

Kaplan, E. Ann. *Trauma Culture: The Politics of Terror and Loss in Media and Literature*. New Brunswick, NJ: Rutgers UP, 2005. Print.

Kavadlo, Jesse. "The Absurd Hero: Escapism, The Dark Knight Trilogy, and the Literature of Struggle." *American Popular Culture in the Era of Terror: Falling Skies, Dark Knights Rising, and Collapsing Cultures.* Santa Barbara, California: Praeger, 2015. 157–186. Print.

Naremore, James. *More than Night: Film Noir in Its Contexts*. Berkeley: U of California, 2009. Print.

Nolan, Christopher et al., "Batman Begins." Directed by Christopher Nolan, Warner Brothers, 2005.

Nolan, Christopher et al., "The Dark Knight." Directed by Christopher Nolan, Warner Brothers, 2008.

Spigel, Lynn. "Entertainment Wars: Television Culture after 9/11." *American Quarterly* 56.2 s (2004): 235-70. Web.

Impact of School Facilities on the Quality of Senior High School Education in China: A Quantitative Study

Chen Zhou

Author background: Chen Zhou grew up in China and currently attends Jiaxiang Foreign Languages School attached to Chengdu No.7 Middle School, located in Chengdu, China. His Pioneer seminar topic was in the field of political economy and titled "Political Participation in the Global World."

I. Abstract

This research attempts to identify the key factors that determine a Chinese senior high school's "quality." To do this, we will compare a group of Chinese senior high schools, including prestigious ones in a typical first-tier city, as well as the average ones in a typical Chinese town. The research will seek major differences in terms of facilities among these selected schools and will quantify them, if possible, to better compare. After gathering the data of these schools, some graphs indicating the correlation between the facilities and the student scores will be made. Conclusions can be drawn from them in turn.

II. Introduction

Research Question: What School-Related Factors Influence Student Outcomes?
Hypotheses
Hypothesis 1: Class size correlates with student outcomes.
Thesis: Larger class size makes better student outcomes due to competition, concentration of teachers, and other factors.
Hypothesis 2: School history correlates with student outcomes.
Thesis: A longer school history makes better student outcomes due to better reputation and more mature management.
Hypothesis 3: Student number and pupil-teacher ratio correlate with student outcomes.
Thesis: Smaller pupil-teacher ratio makes better student outcomes, while student number does not correlate with student outcomes.
Hypothesis 4: Libraries might correlate with student outcomes to some degree, though the correlation is hard to see.
Thesis: The correlation between libraries and student outcomes is hard to examine since virtually every school has a library.
Hypothesis 5: School ownership correlates with student outcomes.
Thesis: Private schools generally have better student outcomes due to better financing.
Hypothesis 6: School location closely correlates with student outcomes.
Thesis: Schools close to cities usually have better student outcomes due to a wide variety of factors.

III. Context

Chinese education has been developing rapidly as a whole since the early 1980s, the time when China began to economically reform. As we can see from Table 1, since 1985, the schools in China have been gradually combined and formalized, while student numbers have kept growing. In the words of China Daily, "this is the transformation of China from a largely illiterate country in 1949, when the People's Republic was established, to a country in which almost all children attend school for nine years and the literacy rate of young people aged 15 to 24 is 99 percent." ("China's Quiet," 2010, para.6). Along with the growth of student numbers, the number of teachers, staff, and workers has been continuously increasing. In response, the educated population has kept expanding, meaning that more people have become educated. Apart from this, the size of the student population grew less markedly each year, but nearly stopped growing in 2006. What can be identified then is that education in China stopped growing in quantity but instead focused on quality. See Table 1. From Table 2, moreover, we can tell that all layers of schools sent more of their students to higher education each year in the period from 1990 to 1999. The rationale behind this might consist of two aspects: both the Chinese government's control and the fact that the schools themselves were continuously becoming better.

UNESCO-IBE once noted, "Educational equity is the prerequisite for social equity" (UNESCO-IBE, 2011). This is true in China as well. However, nowadays in China, despite overall development, educational inequality still commonly exists. Schools in major Chinese cities generally have better ratings than those in average-sized cities or those in the countryside (Journal of Beijing Normal University, 2005). Even in such a superior group, the quality of schools still differs sharply due to various factors.

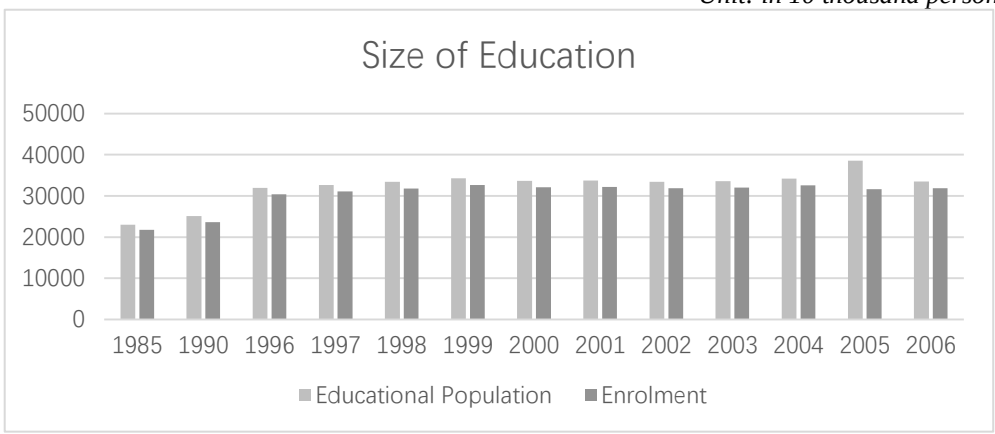
Certain school-related factors determine student outcomes to a very large extent. The study by Andrew Kipnis and Shenfeng Li listed a group of facilities of a "keypoint" high school in China:

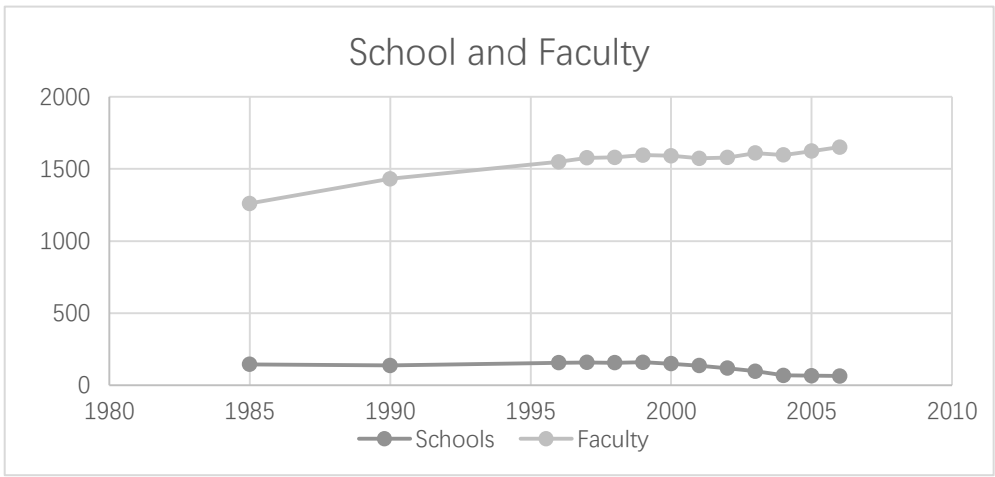
> The facilities are spectacular. All classrooms are equipped with multimedia capabilities; there are computer, science and language labs of many varieties, libraries, reading rooms, acoustically designed theatres for performances by music and theatre students, music and art classrooms with all manner of musical instruments and painting and sculpting equipment, a large outdoor athletics stadium with an Astroturf field and a high impact rubber track, ample basketball, volleyball and ping-pong courts, an indoor sports facility complete with a gymnasium and swimming pool, dormitory space to house all the students with multiple health clinics, pharmacies, can teens and stores scattered among the dorms, and beautiful landscaped grounds". (Andrew Kipnis and Shangfeng Li, 2016, p.7).

However, in schools that are far from cities, things are so different that they directly influence student performance. Zhang & Kanbur described regional and urban-rural differences in both educational provision and attainment as "substantial" (as cited in Hannum & Wang, 2006). This research, in response, attempts to see how much those different factors correlate with student outcomes.

Table 1: *Size of Education*

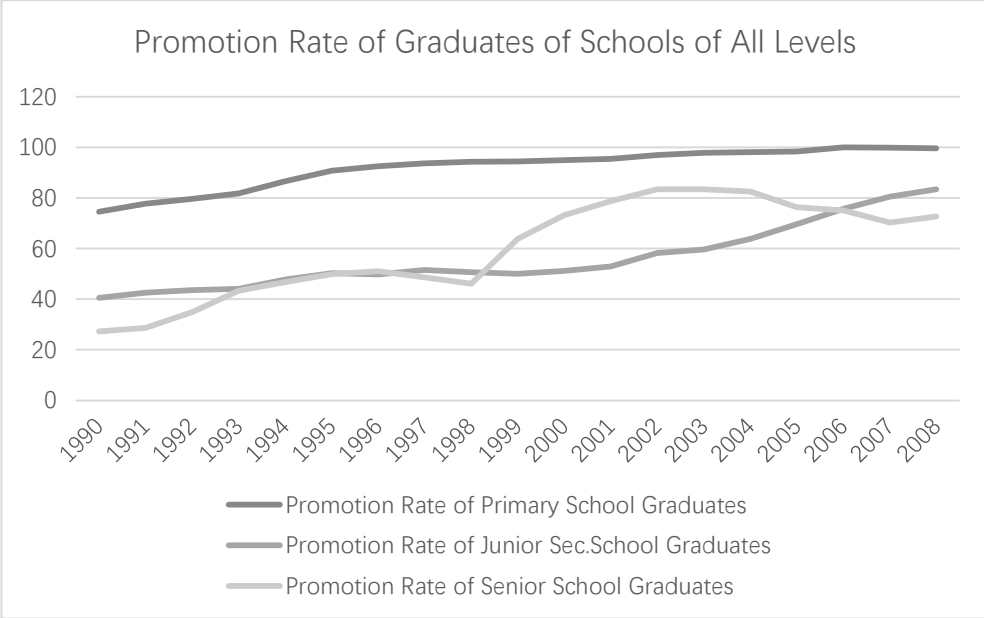
Unit: in 10 thousand person





Adapted from China Education and Research Network,
http://www.edu.cn/gai_kuang_495/20100121/t20100121_442247.shtml

Table 2: *Promotion Rate of Graduates of School of All Levels*



Note: Promotion rate of senior secondary school graduates is the ratio of total number of new entrants admitted to HEIs (including those admitted into the regular full-time courses run by (TRVUs) to the total number of graduates of regular senior secondary schools of the current year.

Adapted from China Education and Research Network,
http://www.edu.cn/gai_kuang_495/20100121/t20100121_441886.shtml

IV. Data and Methods

This research defines "facilities" in a broad sense, which is not limited to physical facilities, like computer labs, experimental labs, multi-media teaching facilities, a good and well-maintained library, all of which are crucial in modern teaching. It would also include other things that are school-related: the teacher-to-student ratio, the class size, and the school's history. Other things that are possibly related to the quality of education are a large and well-maintained playground, a clean and comfortable dormitory (only for boarding schools), a dining hall (some schools do not have one), and classrooms and instruments for the arts (e.g. piano room). Schools will be compared on these facilities. They will not be compared based on a simple "yes-or-no" measurement, but rather on a grading system that can quantitatively examine these factors.

Quality of Education will be measured based primarily upon student Gaokao scores. The Gaokao is a "high-stakes" national college admission exam that determines which school students can go to. Bai, Chi, & Qian explained that the Gaokao "tests students' mastery of the subjects taught in high school." (as cited in Muthanna, 2013, p.2).

However, a simple measure of the average score will not be suggestive. Hence, to diversify our dataset, we are going to incorporate the specific score of all subjects. Though some data is incomplete, there is still a considerable amount of data available.

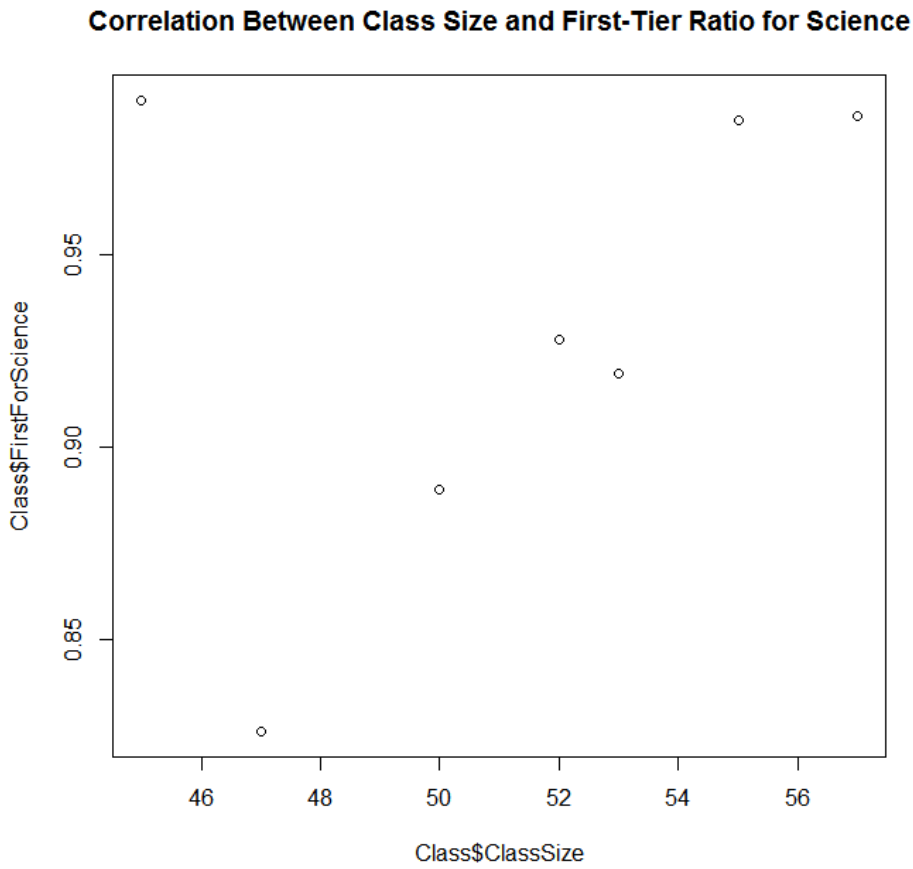
To make our sample representative, we will select a wide variety of schools from both urban and rural areas. The sample will contain both public and private schools.

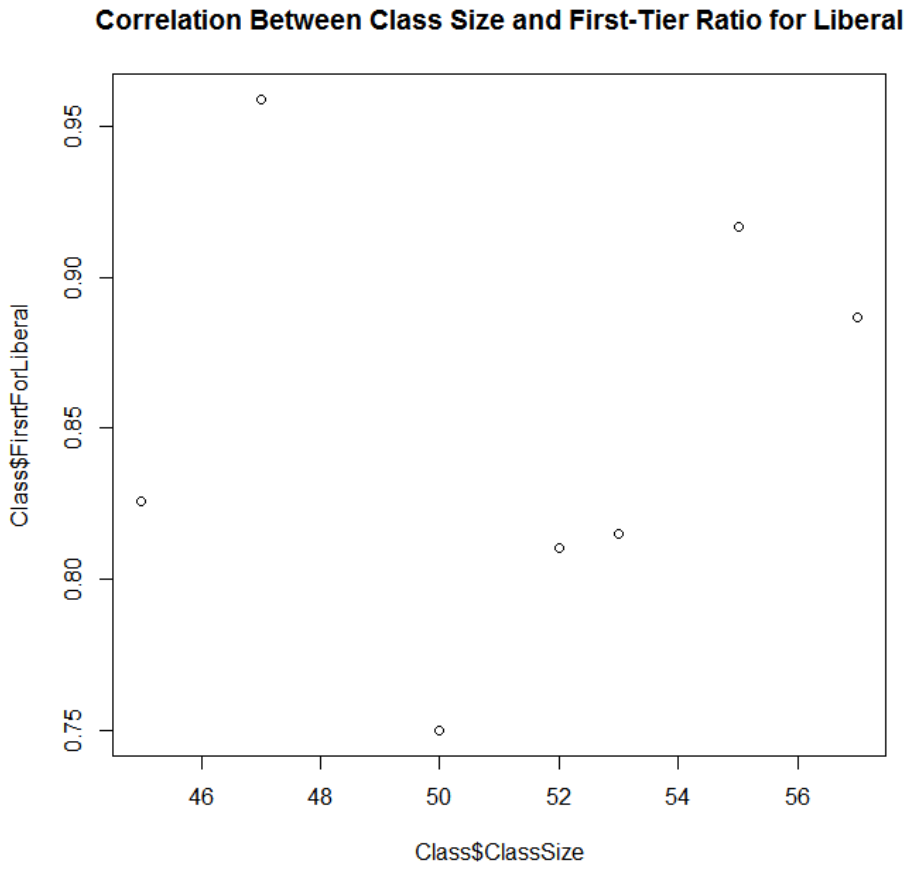
V. Results

5.1 Influence of Class Size

In this section, we will seek the correlation between the class size and students' performance. To eliminate the influence of other factors, we will compare the top schools in the city, since they differ in class size while having most other things in common.

In our dataset, we already have some class size listed. For student performance, we make a graph using school first-tier-ratio for both science and liberal arts. The following are the results





From the first plot above, despite the outlier, a clear correlation pattern can be seen. The exact correlation coefficient is 0.400, which is relatively high.

Therefore, contrary to common sense that larger class size leads to worse scores, our first plot suggests that larger class size actually leads to better scores. A possible reason for this would be that larger class size generates more intensive competition which makes students try harder and gain better scores, or that larger class size can bring more concentrated faculty and improve student scores.

Then we look at the second plot that represents the correlation between class size and first-tier-ratio for liberal arts. The situation is quite different since the correlation pattern is no longer obvious. The exact correlation coefficient is 0.100, a relatively low number.

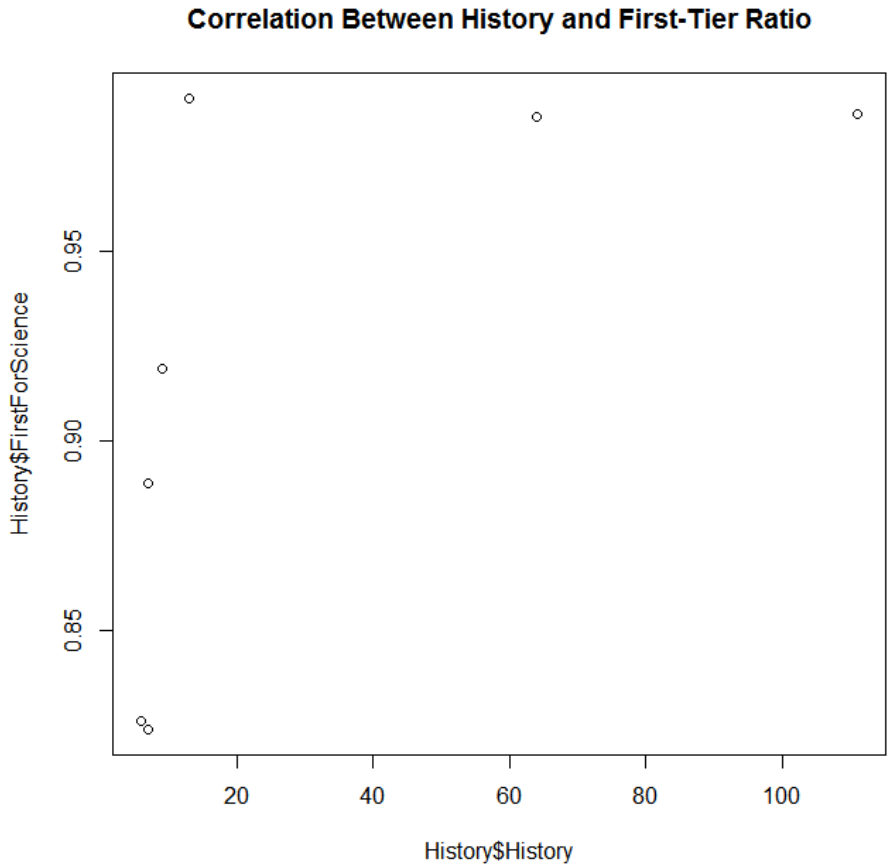
From the observation above, we can tell that as class size increases, the student scores for science actually rise in response, while student scores for liberal arts show weak correlation with class size. Further research needs to be done to clarify the exact reason.

5.2 Influence of History

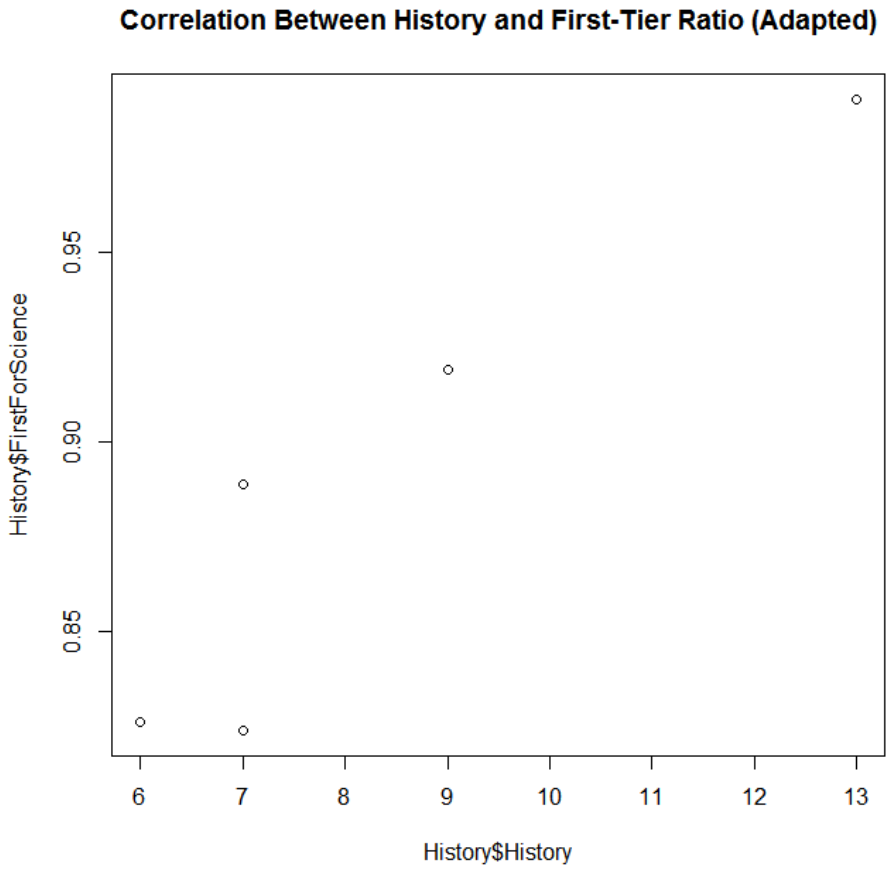
In this section, we will investigate the relationship between when the school was founded and how good the school is. Commonsensically, schools with a longer history would have a stronger reputation and in turn would attract more distinguished students.

To look at this, we picked the founding years of the selected schools to determine how old the schools are. Then we put that together with the schools' first-tier ratio to see the correlation.

An obvious outlier--a school that has over 2000 years of history--has been excluded when making the plot. So the final version is the graph below.



A correlation pattern can be roughly identified. That is, schools with a longer history actually have better student outcomes. The dots that fell within the first twenty years are close to each other. The plot below is a zoomed version that contains only dots between 0 and 20.



From the above, we can see one other fact--as the schools get a longer history, the increase in its first-tier ratio gradually becomes less sharp. This might be due to the fact that schools need a certain time period to establish basic facilities, faculty, and reputation. After that growth period, the schools themselves will grow less and less as time passes.

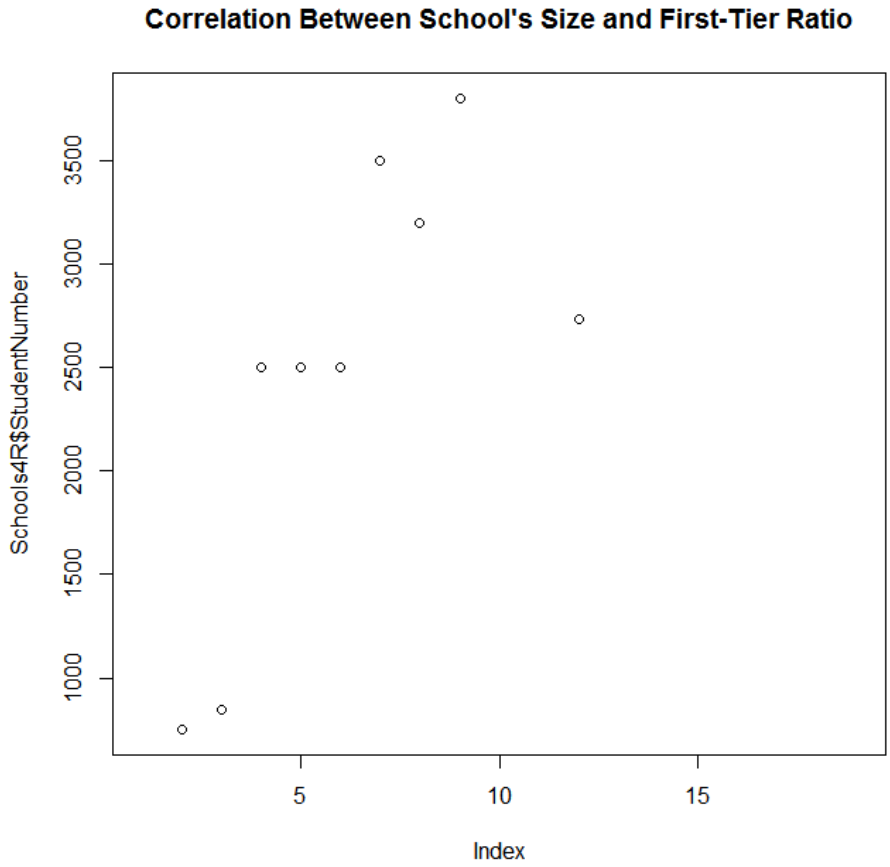
5.3 Influence of Student Number and Pupil-Teacher Ratio

Another key determiner of a school's quality might be related to people. To properly consider at this, we will look at these data separately.

5.3.1 Student Number

Student number may influence school performance to some extent, though in China, some good schools have a large student body on account of their good reputation. That is, they usually attract a lot of students, or otherwise the government would require them to expand their campus to hold more students in order to improve the local education quality.

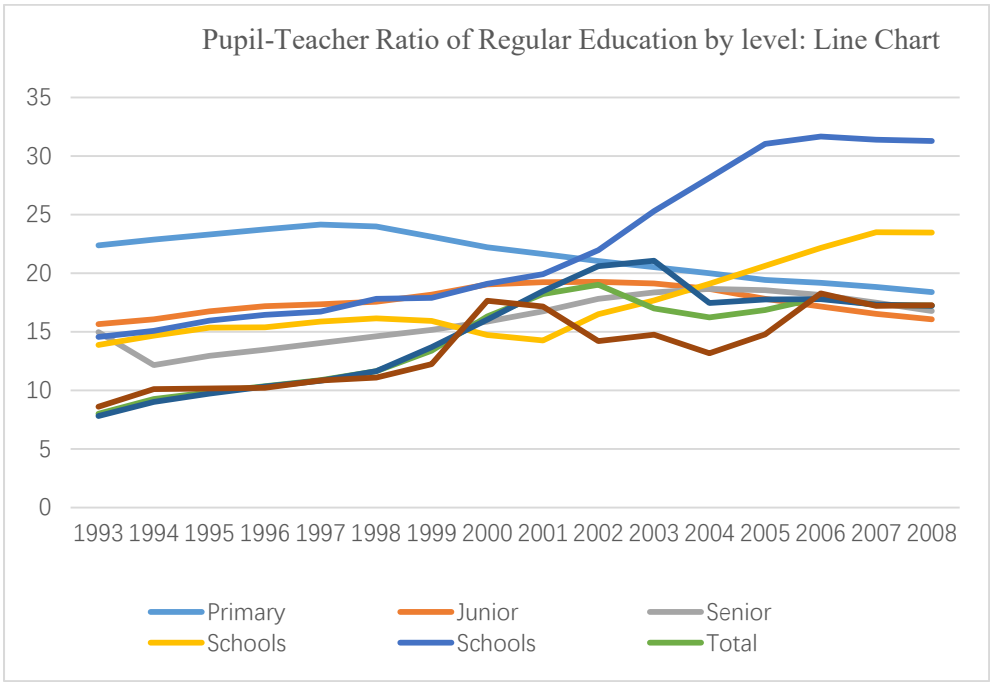
The statistics would probably suggest something other than the above-mentioned. Here is the plot we got after correlating student number with the student performance.

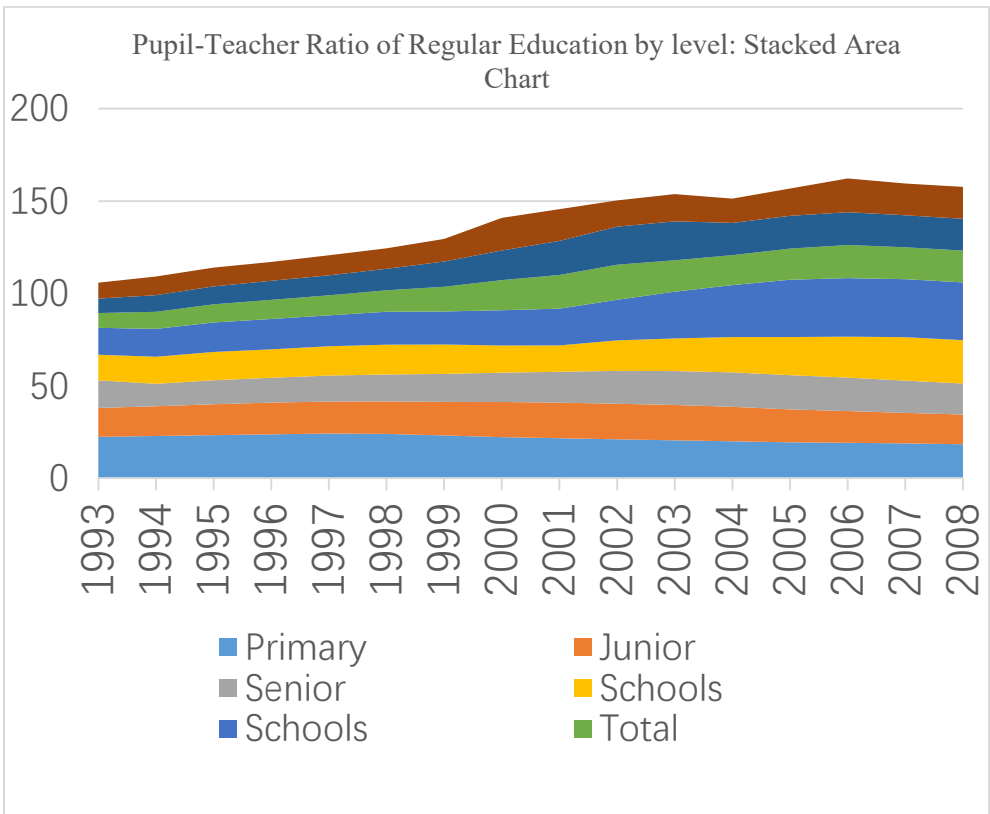


The correlation coefficient is 0.008. Therefore, the correlation pattern is not obvious in the graph above. So, the school's size probably does not influence student outcomes as much as pupil-teacher ratio does.

5.3.2 Pupil-Teacher Ratio

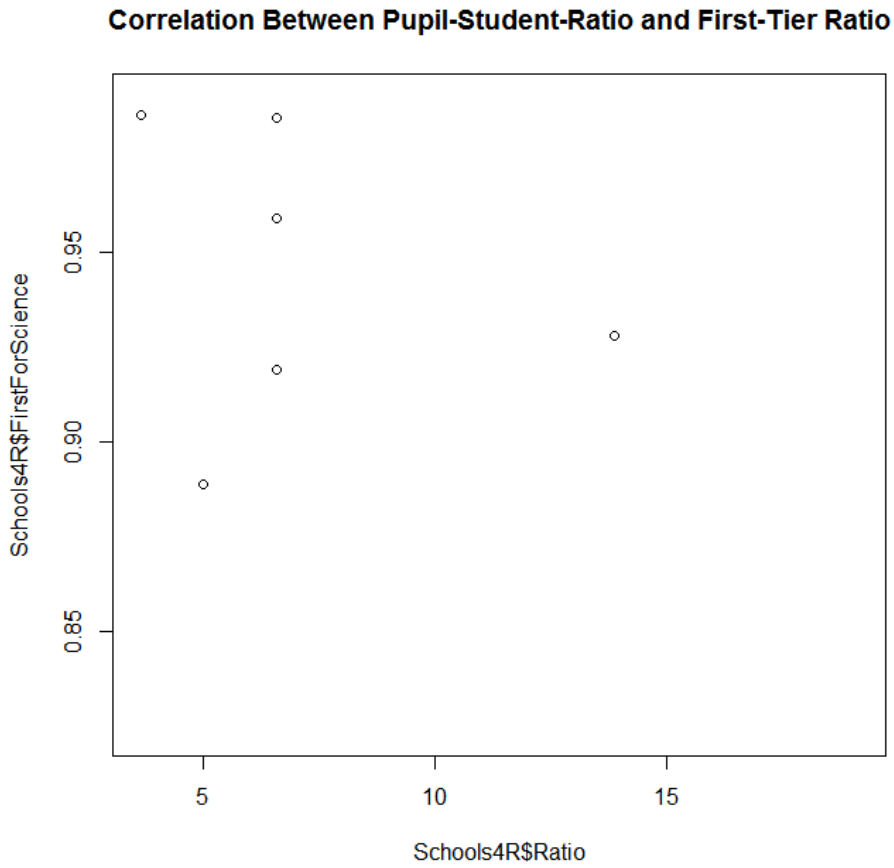
It would seem that pupil-teacher ratio overlaps with the student number in meaning and that there is no need to do both. However, if we look at it carefully, it serves as a different indicator than student number. A school that has a large student body would probably have a large pupil-teacher ratio as well. If this is the case, the school, by common standards, can still be defined as a "good" school since every student is sufficiently covered by faculty. However, in other cases, some large schools do not have a good pupil-teacher ratio. In these cases, the school might look crowded and might not have enough faculty to cover every student sufficiently.


Pupil-Teacher Ratio of Regular Education by level: Line Chart
35
30
25
20
15
10
5
0
1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
Primary Schools
Junior Schools
Senior Schools
Total


Pupil-Teacher Ratio of Regular Education by level: Stacked Area Chart
200
150
100
50
0
1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
Primary
Junior
Senior
Schools
Schools
Total

The dataset we have attest to the average pupil-teacher ratio in China as the charts above suggest. It shows a trend of constant growth over time until 2006. Then the pupil-teacher ratio stopped rapid growth.

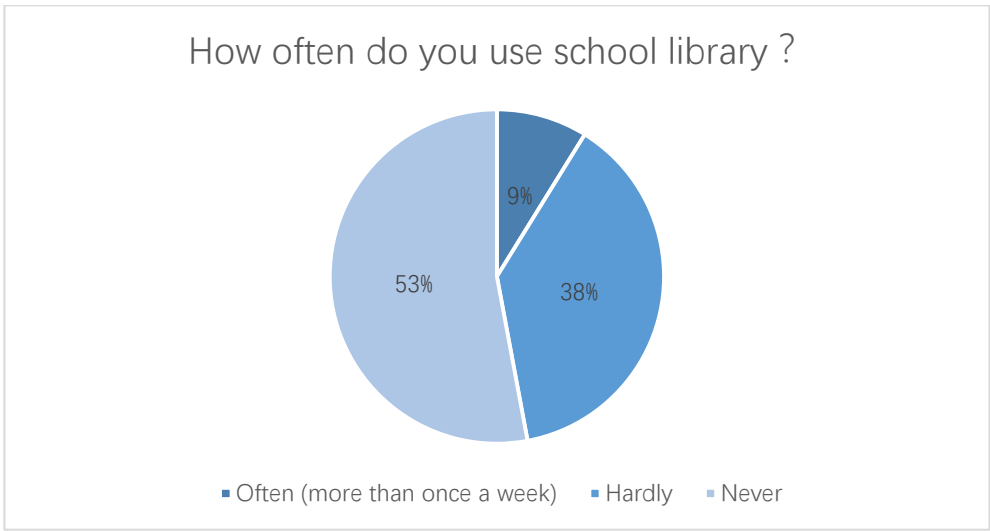
Now, with the plot below, we can see how much this factor actually weighs.

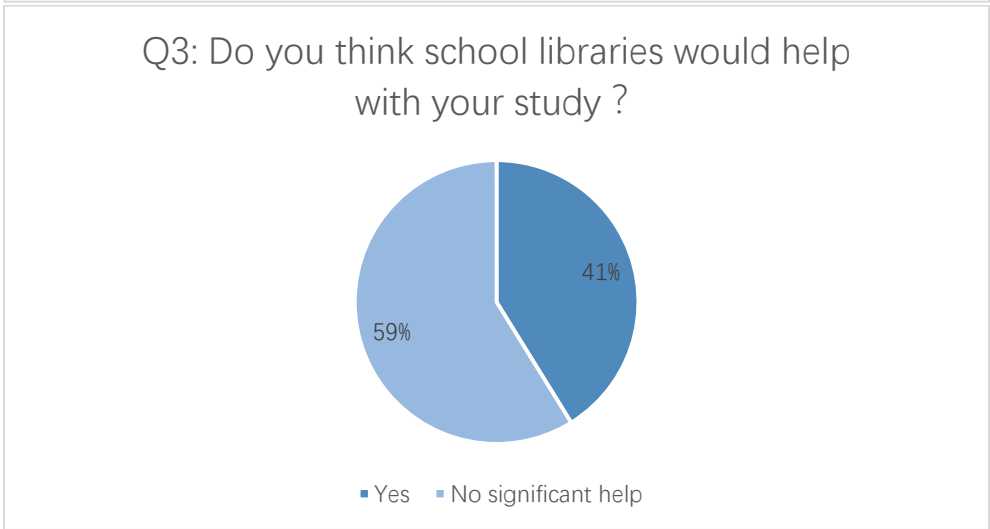


As we can see above, though not obvious, as the pupil-student ratio grows, the first-tier ratio--an indicator of student outcomes--gradually decreases. This would suggest that pupil-teacher ratio does have an effect on student performance, and that students better covered by teachers demonstrate better performance.

5.4 Influence of Library

Libraries, particularly libraries with a wide variety of books, may have a positive impact on students. Students can normally do research in the library and enrich their knowledge. In China, however, the situation is not exactly the same. Students typically do not use the libraries much in high schools. In most cases, even if they use the library sometimes, they do not use it to do academic research. Recreational use of libraries is common in Chinese high schools. That is, students use libraries to read magazines, comics, etc., which does not have a substantial positive influence on their scores. In our dataset, most schools selected (except one) have libraries. Therefore, analyzing such data would be unnecessary. Instead, a survey done recently will be drawn here. It surveys a considerable number of students from several typical schools. Their habits of using libraries and their attitudes toward them are indicated below.





Raw data is in the appendix.

5.5 Influence of School Ownership

School ownership would have some influence on student performance. Private schools are normally better funded than public schools since they charge higher tuition. Private schools are also independent from government planning, which will enable them to select students on their own. In this way, the students enrolled in private schools may have a better starting point, and in turn, do better. In other cases, however, public schools are better than private schools since they are so called "key schools." Such "key schools" are typically well-funded and have a large reputation.

School Name	First for Liberal	First for Science	First-Tier-Ratio	Private
Jiaxiang	0.826	0.9903	0.9603	√
JXPixian	0.75	0.889		√
Shude Foreign	0.959	0.826	0.919	√
JX Chenghua	0.818	0.824		√
No.7 High	0.8867	0.9864		
Shude Ningxia	0.917	0.9853		
Shude Guanghua	0.815	0.919		
Shishi Wenmiao	0.8103	0.9279	0.9076	
Shishi Beihu				
Liewu High School			0.65	
Yandaojie			0.47	
Railway			0.632	
Luojiang High School			0.3333	

5.6 Influence of School Location

Besides the factors listed above, school location is another important factor, which influences student outcomes to a very large extent. As we can see from our dataset, the two schools from a town have far lower performance scores than schools in the city, even though they have all the facilities schools in cities have.

Decentralized education finance schemes that require local governments to provide the majority of funds for schools may be associated with greater inequality. (E. Hannum M. Wang, 2006). Because of lack of funding and elite students and teachers, schools in the countryside do not typically gain high ratings. Such situations are so common in China that many rich families in towns or villages send their children to schools in cities.

School Name	First for Liberal	First for Science	First-Tier-Ratio	City
Jiaxiang	0.826	0.9903	0.9603	√
JXPixian	0.75	0.889		√
Shude Foreign	0.959	0.826	0.919	√
JX Chenghua	0.818	0.824		√
No.7 High	0.8867	0.9864		√
Shude Ningxia	0.917	0.9853		√
Shude Guanghua	0.815	0.919		√
Shishi Wenmiao	0.8103	0.9279	0.9076	√
Shishi Beihu				√
Liewu High School			0.65	√
Yandaojie			0.47	√
Railway			0.632	√
Luojiang High School			0.3333	
Canting High School			0.15 (estimated)	

VI. Discussion and Conclusions

This paper has examined a group of school-related factors that potentially influence the quality of education in Chinese senior high school, and has come up with some interesting findings.

First, as opposed to common sense, bigger class size actually leads to better student performance, particularly in top schools in cities. This might result from more intensive competition and better concentration of teachers. Second, schools that were founded earlier in time generally have better student outcomes, on account of their reputation and their finished growing periods. Third, size of student body does not influence student performance much, while pupil-teacher ratio does, when every student is covered more sufficiently by faculty. Fourth, libraries have inconsiderable impacts on student outcomes, since students use them for mostly recreational purposes. Fifth, private schools generally have better student outcomes than public schools because they are usually better funded. Sixth, schools in cities have better student performance than schools out of cities, except in some extreme cases.

The research only draws a part of the picture due to the limitedness of the samples. Some other key determiners still have not been researched: school admission rates, certifications of teachers, and school financial conditions. These factors might be paramount, since they are closely related to the daily operation of schools. Beside this, it is possible that problems that have some influence on student outcomes might still have not been seen. For them, it would be advisable to delve deep into the schools, watch closely how they work and the problems they might have, and finally, draw conclusions.

Appendix: Dataset (Omitted for Publication)

Appendix: Survey

Question 1: How often do you use school libraries?

Answer Choices	Number
Often (more than once a week)	3
Hardly	13
Never	18

Question 2:
Invalid due to the limited sample size.

Question 3: Do you think school libraries help with your study?

Answer Choices	Number
Yes	14
No significant help	20
Sample Size	34

References:

Bai, C., Chi, W., & Qian, X. (2014). Do college entrance examination scores predict undergraduate GPAs? A tale of two universities. China Economic Review, 30, 632-674. http://dx.doi.org/10.1016/j.chieco.2013.08.005

HANNUM, E., & WANG, M. (2006). Geography and educational inequality in china. China Economic Review, 17(3), 253-265. doi:10.1016/j.chieco.2006.04.003

(2005). To narrow down the gap:a momentous issue in chinese education policy. Journal of Beijing Normal University.

Kipnis, A., & Li, S. (2010). Is chinese education underfunded? The China Quarterly, 202(202), 327-343. doi:10.1017/S0305741010000263

Lin, Z. (Ed.). (2010, April 04). China's Quiet Education Revolution. China Daily. Retrieved September 15, 2016, from http://news.xinhuanet.com/english2010/indepth/2010-04/21/c_13260769.htm

Muthanna, A., & Sang, G. (2015). Undergraduate chinese students' perspectives on gaokao examination: Strengths, weaknesses, and implications. International Journal of Research Studies in Education, 5(2), 3-12. doi:10.5861/ijrse.2015.1224

U. (2010, November). World Data on Education: People's Republic of China. Retrieved September 14, 2016, from http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/China.pdf

Zhang, Xiaobo, & Kanbur, Ravi (2005). Spatial inequality in education and health care in China. China Economic Review, 16, 189—204.

Bibliography:

Bai, C., Chi, W., & Qian, X. (2014). Do college entrance examination scores predict undergraduate GPAs? A tale of two universities. China Economic Review, 30, 632-674.http://dx.doi.org/10.1016/j.chieco.2013.08.005

Golley, J., & Kong, S. T. (2016). Inequality of opportunity in china's educational outcomes. China Economic Review, doi:10.1016/j.chieco.2016.07.002

HANNUM, E., & WANG, M. (2006). Geography and educational inequality in china. China Economic Review, 17(3), 253-265. doi:10.1016/j.chieco.2006.04.003
(2005). To narrow down the gap:a momentous issue in chinese education policy. Journal of Beijing Normal University.

Kipnis, A., & Li, S. (2010). Is chinese education underfunded? The China Quarterly, 202(202), 327-343. doi:10.1017/S0305741010000263

Tang, J. (2016). "Lost at the starting line": A reconsideration of educational inequality in china, 1978–2008. The Journal of Chinese Sociology, 3(1), 1-18. doi:10.1186/s40711-016-0028-z

U. (2010, November). World Data on Education: People's Republic of China. Retrieved September 14, 2016, from http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/China.pdf

Wu, Y. (2008). Cultural capital, the state, and educational inequality in china, 19491996r. Sociological Perspectives, 51(1), 201. doi:10.1525/sop.2008.51.1.201

The Neural and Cognitive Basis of Dreaming

Ria Tomar

Author background: Ria Tomar grew up in the United States and currently attends Mission San Jose High School, located in Fremont, California. Her Pioneer seminar topic was in the field of neuroscience and titled "The Decision Making Brain"

I. Introduction

1.1 Abstract: Purpose of the Research

As humans, we live for our dreams and aspirations, yet we know so little about their biological purpose and function. The neural basis of dreaming is an enigma that many scientists have pondered for centuries. From Plato and Aristotle's theory of consciousness to Sigmund Freud's hypothesis that dreaming is essentially a safety valve for our unconscious desires, numerous theories have made their way into the medical journals. Today, however, scientists tend to generally support one of these two surprisingly contrasting theories: activation synthesis and threat simulation. Activation synthesis states that dreams result from activation during a certain stage of sleep, due to excess electrical impulses. (8,22) However, due to a multitude of research papers and scientific discoveries, it is a well-known fact that our human body regulates processes for a reason and is extremely efficient in the process. The Activation Synthesis Hypothesis implies that dreaming is a result of excess material, which according to the general rules of nature and science is not the likely case. On the other hand, the threat simulation theory describes that dream consciousness is essentially an ancient biological defense mechanism, evolutionarily selected for its capacity to repeatedly simulate threatening events. (23) This theory is more agreeable with nature's tendencies because it suggests that dreaming occurs for a definitive purpose and is not simply a result of excess electrical impulses. Although the threat simulation theory seems to be a well-suggested and simplistic answer to the mystery that surrounds dreams, this theory does not account for the dreams that are bizarre, related to memories, or in the form of a narrative.

II. Purpose

2.1 The Model Dream Theory

Taken from the model dream theory, there are three main questions that need to be answered in order to fully understand dreaming and its purpose:

- How does neurophysiology set the stage for memory selection?
- How do dreams favor associative processes that produce bizarre and sometimes unrecognizable representations of memories?
- How do those elements combine with others in a narrative with high emotional content? (21, 22, 24)

2.2 Hypothesis

Taken from the model dream theory, there are three main questions that need to be answered in order to fully understand dreaming and its purpose:

- How does neurophysiology set the stage for memory selection?
- How do dreams favor associative processes that produce bizarre and sometimes unrecognizable representations of memories?
- How do those elements combine with others in a narrative with high emotional content? (21, 22, 24)

2.3 The Sleep Cycle and Dreaming

Since dreaming is a process that occurs during the phenomenon that is sleep, it is important to pinpoint and examine all the different stages of sleep. In order to properly study dreams it is important to research where they originate. Sleep is composed of four main stages:

Figure 1.1 *Sleep Cycle Breakdown (32)*

- Stage 1
  - During stage 1, an organism is essentially easing into sleep. (1)
- Stage 2

- Stage two is when eye movements come to a stop and your brain waves presented in an EEG (electroencephalogram) begin to slow. (4) The time elapsed until the end of stage 2 tends to be about an hour. It is not known if dreaming occurs in these first two stages.

- REM Sleep
  - REM sleep is essentially electrical activity in the brain that mimics the brain activity when an organism is awake. (10) In order to keep a "resting body" asleep and completely halt muscle activity, the brain releases hormones such as norepinephrine, histamine, and serotonin. (13) This stage of sleep is a phenomenon at the behavioral level. Brain activity in REM sleep is drastically different than brain activity in stages 1 and 2. During REM sleep, the category of dreaming that occurs is called active dreaming. Slow brain waves called delta waves are interspersed with fast brain waves that result in the phenomenon of rapid eye movement. This stage tends to occur 70-90 minutes after we fall asleep. (1)
  - During NREM (non-rapid eye movement) sleep, the muscles are extremely relaxed and do not seem to be moving. It is important to note that the muscles aren't completely "paralyzed by the body." (12) For example, in REM sleep, the brain utilizes inhibitory signals, using the neurotransmitter norepinephrine to control muscle movement and completely suppress movement of every part of the body that isn't absolutely necessary for survival:
    - Eye movements
    - Heart
    - Diaphragm and one or two other essential functions, that allow us to breathe and remain alive (15)
  - This suppression of muscle movement in the body during the REM stage of sleep is called atonic. (15) This process is guided and started by the pons region of the brain stem called locus couruleus. (15)

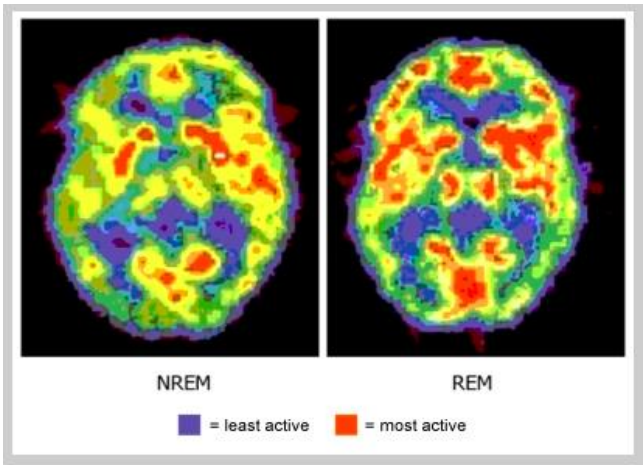


Figure 1.2 *Depiction of brain activity in two different stages of sleep (15)*

  - Also, the image below presents a very important point. The brain activity during REM sleep is extremely similar to our brain activity when we are awake, except that the REM sleep waves seem to be just a faster and more clumped together version of the EEG brain waves under Awake. (24)

- Stage 4

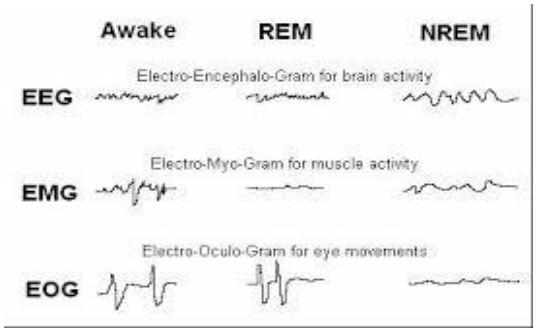


Figure 1.3 *Shows three different stages and their respective brain waves (30)*

○ The final and most mysterious stage of sleep is stage 4, also known as "deep sleep." (1) In this stage there are no sudden eye or muscle movements. (2) Brain activity during this stage is mostly composed of theta waves, alpha waves and even the high frequency beta waves more typical of high-level active concentration and thinking, as shown in the picture above. (1) On an electroencephalogram these waves show up in a very rigid and "saw-tooth" like manner. (14) Even if we do not know everything about sleep, we know it is a necessary process without which our body will fail.

III. Neuromechanisms that relate to dreaming

3.1 Where do Dreams Occur?

This section aims to discuss physical and psychological location of dreaming.

Dreaming and REM Sleep

Dreaming can be seen as a "state of delirium possessing qualities of hallucination, disorientation, confabulation, and amnesia." (12) Once awakened from REM sleep, around 70-95% of people can remember their dreams. However, during NREM sleep this number reduces to 5-10%. (12)

Different neurotransmitters are released in order to activate different stages of sleep. Serotonin, an inhibitory neurotransmitter, constricts certain muscles and triggers NREM during the onset of sleep. (18) NREM then switches to REM sleep in the pons, located at the base of the brain, when acetylcholine is released. However, the re-release of noradrenaline and serotonin switches off REM sleep and reactivates NREM sleep again. This system of activation and inactivation of these stages of sleep is known as reciprocal interaction.(12) After acetylcholine is excreted, the pons sends out signals to the thalamus, which is then relayed to the cortex. Also, neurons in the spinal cord are "shut off," which creates temporary paralysis of the body.

Dreams originate in the frontal lobe and the back of the brain. This suggests that REM sleep and dreaming are very much related.

- 26 cases relating to damage to the pons caused loss of REM sleep
- Ability to dream remained intact for 25/26 patients

(12) This also suggests that if the frontal areas of the cortex are damaged and the ability to dream is diminished, that does not necessarily mean that the REM cycle will be tarnished as well. (12) However, it is important to note that REM sleep can be seen as one of the triggers

for dreaming. In order to safely go through the whole process of dreaming, the pons paralyzing us dreamers prevents us from acting out our dream.
The frontal lobes are responsible for dreaming and they have a "large fiber pathway, which transmits the neurotransmitter dopamine from the middle part of the brain to the higher parts of the brain." (12)

An increase of dopamine stimulants, L-dopa, greatly intensifies the amount and frequency of dreams. (12,13) The frontal and limbic areas of the brain also happen to be concerned with arousal, memory, emotion, and motivation.

Sandman Switch

Recently, scientist Diego Pimentel has discovered an ion channel named Sandman that is related to the "sleep homeostat," which controls when we go to sleep. This ion channel is kept inside the sleep control neurons when electrically active. Interestingly enough, when dopamine is present, it causes Sandman to move outside of the cell, which short circuits the neuron and shuts them off. This causes wakefulness. The discovery of the Sandman Switch is a huge step forward when it comes to controlling sleep. This switch is like a thermostat, but instead of temperature it responds to a neurotransmitter called dopamine.

Dopamine is a neurotransmitter that is associated with pleasurable reward, movement, memory, focus, sleep, and learning. (13) The recent discovery that levels of dopamine in the brain regulate sleep shows the scientific world the possible functions of sleep. Using positive feedback mechanisms, the sandman switch utilizes dopamine to control our "sleep homeostat."When we have low levels of dopamine, or lack of will to move, learn, focus, or remember, our body decides to shut down and refresh itself through the process of sleep. As dopamine levels rise, our body shuts off the "sleep control neurons" and we are supposedly entering a new day with more ambition and drive to face reality.

3.2 Who can dream?

The first step to answering these questions is establishing what organisms actually dream. Since we cannot physically study dreams in other organisms we must research REM (rapid eye movement) sleep due to its direct relation to the process of dreaming. REM sleep is temperature regulated; (18) hence, in order to reach REM sleep, an organism must get its body to a certain temperature. Warm-blooded animals can reach this temperature, hence REM sleep, whereas cold-blooded animals cannot reach that temperature (find the temperature). Warm-blooded animals are more evolutionarily advanced than cold-blooded animals, which suggest that REM sleep, and, in turn, dreaming is an evolutionary advantage that is a necessary component of intelligent life.

3.3 Neurophysiology and Memory Selection

Although many different studies have been conducted that investigate how sleep contributes to remembering or forgetting memories, very few studies have actually focused on the relationship between dream content and memory consolidation in sleep. (24)
The evidence from the minimal studies conducted suggests that people have to be sorting through memories while they are dreaming. They have to be choosing to remember some memories and choosing to forget others, based on what is needed for survival. (23, 24)
Since most researchers suggest that dreams are and always have been a conscious experience that occurs during sleep, it is essential that a contradicting theory be presented in order to fully understand the whole memory process during sleep. Dennett's cassette theory of dreaming suggests that dreams are instantaneous memory insertions that occur at the

moment of awakening. In “Are dreams experiences?” Dennett separates dreaming into two main processes: (28)

- Composition
  - Responsible for the content of certain narratives and stories we dream about during sleep
- Memory Loading
  - Ability to remember after waking up from sleep

Most scientists completely understand Dennett’s logic up to this point of his research. As soon as Dennett stops maintaining sleep consciousness as an essential prerequisite for dreaming, these very scientists choose to support the better-known theory of threat-simulation and memory processing. It is important to note that both theories provide ample evidence for their respective theories of the relationship between dreaming and REM sleep. Also, since the process of dream recalling is extremely unreliable on a subject-to-subject basis, it is very difficult to distinguish between the two contrasting theories. (28)

One infamous dream reported by Maury in 1861 reinforces Dennett’s cassette theory.

The basic idea behind Dennett’s cassette theory goes back to a famous dream reported by Maury (1861), in which a long and complex dream about the French revolution culminated in his execution on the guillotine. At this point, Maury awoke to find that the headboard of his bed had fallen on his neck. Because the dream seemed to systematically build up to its dramatic climax, which in turn was caused by an external stimulus, he and others suggested that such cases were best explained as instantaneous memory insertions experienced at the moment of awakening. Many dream researchers discussed this theory, also known as the Goblot-hypothesis. (28)

Certain elements of a dream often are derived from waking events. However, it is important to note that just because an element can be traced to a specific event does not mean that memory was used for dream construction. The memories associated with dreaming are called episodic memories. “Episodic memories are defined as a memory of an event, recalled as an integrated whole, with the actual waking event replayed in one's mind.” (21). Dream researchers tend to also ask if the dream element source is from a waking event. This is to test whether the process is independent of how memories are actually stored in the brain.

Surprisingly, when subjects of research choose to identify waking events as dream elements, episodic memories are rarely replayed by dreams. In fact, 364 dream elements out of 299 dream reports were analyzed, and it was revealed that a very small percentage, around 1 - 2%, had properties of episodic memories. (21)

This evidence suggests that dreams may be traces of certain episodic memories and components, rather than episodes that occur solely based on our experience while we are awake.

Any theory relating to the function of dreaming has to be able to explain why certain dreams are forgotten rather than remembered and vice versa. It has been famously proposed by scientists that REM sleep may play a significant role using “dreaming” as a process for erasing and storing memories. (21) If this were true, it would prove why we could only remember a handful of dreams. Also, this theory would suggest that remembering dreams would actually undermine this memory processing system. However, before generalizing and solidifying this process in all types of dreams, it is important to classify the different types of dreaming and how the process works in each of them. In other words, which cognitive functions are independent of dreaming and which aren’t?

IV. The Cognitive Aspect of Dreaming

4.1 Drawbacks of the Threat Simulation Theory

TABLE 1. Frequency and Gender Differences of Typical Dreams (in Percentage)

Item	Total ($N = 444$)	Women ($n = 376$)	Men ($n = 68$)	Log. Reg.[a] χ^2	p
31. School, teachers, studying	89.2	89.4	88.2	0.1	.8019
1. Being chased or pursued	88.7	89.1	86.8	0.1	.7767
32. Sexual experiences	86.7	85.6	92.7	4.0	.0455
12. Falling	74.3	74.2	75.0	0.4	.5471
6. Arriving too late	68.5	70.0	60.3	2.0	.1587
36. A person now alive being dead	68.0	70.7	52.9	5.8	.0160
11. Flying or soaring through the air	63.5	63.0	66.2	0.9	.3437
38. Failing an examination	60.8	63.8	44.1	7.9	.0051
37. Being on the verge of falling	56.5	57.7	50.0	0.8	.3655
4. Being frozen with fright	56.3	56.7	54.4	0.0	.9518
35. A person now dead being alive	45.0	46.3	38.2	1.6	.1993
2. Being physically attacked	44.8	44.7	45.6	0.6	.4485
14. Being nude	43.0	43.1	42.7	0.0	.8277
5. Eating delicious food	42.1	43.6	33.8	1.0	.3072
7. Swimming	38.7	38.0	42.7	1.4	.2429
8. Being locked up	38.7	37.2	47.1	3.3	.0674
24. Insects or spiders	37.2	37.0	38.2	0.8	.3732
27. Being killed	36.3	34.8	44.1	5.9	.0155
18. Your teeth falling out/ losing your teeth	35.6	35.6	35.3	0.1	.7428
15. Being tied, unable to move	34.7	33.5	41.2	2.4	.1213
13. Being inappropriately dressed	33.1	33.2	32.3	0.0	.8370
50. Being a child again	32.7	32.7	32.4	0.0	.8430
3. Trying again and again to do something	30.4	29.5	35.3	0.6	.4547
30. Being unable to find, or embarrassed about using a toilet	30.0	31.1	23.5	1.0	.3278
53. Discovering a new room at home	29.1	29.0	29.4	0.2	.6344

TABLE 2. Frequency and Gender Differences of Typical Dreams (in Percentage)

Item	Total ($N = 444$)	Women ($n = 376$)	Men ($n = 68$)	Log. Reg.[a] χ^2	p
45. Seeing a face very close to you	26.1	26.9	22.1	0.3	.5691
9. Snakes	25.0	25.8	20.6	0.6	.4314
20. Having magical powers	23.6	23.4	25.0	0.3	.5984
29. Vividly sensing, but not necessarily seeing or hearing, a presence in the room	23.6	23.7	23.5	0.0	.9383
10. Finding money	19.8	17.6	32.4	8.5	.0036
21. Floods or tidal waves	19.8	20.0	19.1	0.1	.7803
42. Killing someone	18.7	16.2	32.4	12.1	.0005
28. Seeing yourself as dead	18.2	17.8	20.6	0.7	.3933
44. Being half awake and paralyzed in bed	18.0	17.0	23.5	2.4	.1523
43. Lunatics or insane people	17.1	17.0	17.7	0.2	.6564
19. Seeing yourself in a mirror	16.9	18.1	10.3	2.0	.1530
25. Being a member of the opposite sex	16.0	17.8	5.9	4.5	.0335
39. Being smothered, unable to breathe	16.0	17.0	10.3	1.1	.3021
52. Encountering god in some form	12.4	11.7	16.2	0.9	.3493
54. Seeing a flying object crash	11.9	10.9	17.7	3.2	.0730
23. Earthquakes	11.7	12.0	10.3	0.2	.6442
51. Seeing an angel	11.3	11.2	11.8	0.0	.9320
17. Creatures, part animal, part human	11.0	10.4	14.7	0.7	.4068
22. Tornadoes or strong winds	7.9	8.0	7.4	0.0	.8940
41. Being at a movie	7.7	7.7	7.4	0.0	.9148
47. Seeing extra-terrestrials	6.5	6.4	7.4	0.3	.5674
48. Traveling to another planet	6.3	6.1	7.4	0.3	.6166
49. Being an animal	5.0	4.0	10.3	5.3	.0209
46. Seeing a UFO	4.5	3.5	10.3	6.8	.0090
55. Someone having an abortion	2.9	3.2	1.5	0.0	.9668
26. Being an object	2.5	2.1	4.4	0.6	.4463

[a]Logistic regression with the variables of gender (depicted), age, and number of recalled dreams.

One study done in 2004 "investigated the stability of the rank order of the dream themes and of gender differences in the content of dreams." It was clear that this paper was clearly in favor of the theory of threat simulation and the evidence definitely supported it.

***Figure 2.1** The evidence from the study conducted is placed below (27)*

This data from the study told the researchers that each person has a mean value of 17-18 themes of dreaming and around ten dreams per month. Based on the evidence, we can also conclude that women and men, based on their respective social standards and fears, dream about very gender-specific fears. This study states that we can divide dreams into three main clusters. This paper fails to mention why any cluster besides the one that poses direct threat to humanity exists. Threat simulation seems to be on the right track by suggesting that dreaming can be for survival, but this theory does not account for all the dreams that have no realistic survival purpose. Yes, threat simulation is a valid theory, but it is not an all-encompassing theory. Flying and eating delicious food, as listed in the charts above, are not threats to human existence on even the most privileged level. It is clear that there has to be another reason why dreams occur. (27)

4.2 Emotions and Dreams

What differentiates dreams and nightmares, the two main categories of dreams, are emotions. The raw emotions of fear, anxiety, happiness, excitement, and the like, drive the narrative created in dreams. During REM sleep, the amygdala, which is known as the emotional center of the brain, activates the medial prefrontal cortical structures that regulate the highest order of emotions. DLPFC, dorsolateral prefrontal cortex (shown in Figure 2.2), is also associated with higher cognitive functions and this also may add a certain amount of emotional processing to dreaming. Through creating narratives that are associated with emotions, the brain seems to be relatively biased towards emotional processing in the REM state.

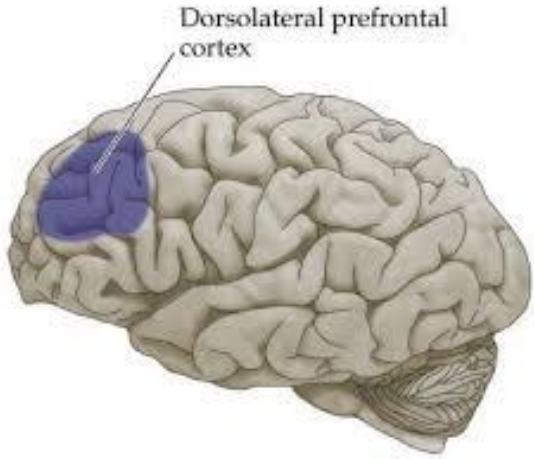


Figure 2.2 *Demonstrates location of Dorsolateral prefrontal cortex (31)*

Evidence for emotion enhancement in REM sleep exists. (15) Also, the viewing of negative emotion, whether it is from movies, anxiety, or depression, correlates with certain reports of negative dreams or nightmares. (20) Obviously, most dream research must rely on its subjects, and unfortunately the ability to deduce dreams and emotions without receiving skewed results is extremely difficult for the scientist. However, sleep onset dreaming may be able to simplify that process and provide more accurate results

4.3 Sleep Onset Dreaming

Hypnagogic dreams, dreams that occur between the state of lucid and deep sleep, or sleep onset, can help eliminate a lot of the problems that come with discovering sources of dreams. (21, 28) Hypnagogic dreams are experimentally controllable and are more likely to include daytime experiences. (28)

Here is an example of an experiment conducted by researchers using sleep onset as a tool to monitor dreams and the effect of different video games on them:

"The researchers had manipulated hypnagogic dream content by having subjects play the video game Tetris or the arcade style downhill skiing simulator Alpine Racer II. Reports were collected from subjects in their own homes, with their sleep monitored by the Nightcap sleep monitoring system, rather than standard polysomnography. Using these games, they obtained sleep onset reports of images of Tetris or downhill skiing in up to 89% of subjects and 42% of first-night reports, with no difference in frequency or content between normal and densely amnesic subjects. Nevertheless, the neocortical sources of these images were not simply stored sensory representations of recent stimuli, as Tetris players occasionally reported images from past versions of Tetris and Alpine Racers reported images from actual skiing." (21)

Based on these experiments and studies, it can be concluded that a very high rate of memories from previous events during the day or older related memories play a huge role in hypnagogic dreaming without involvement from the medial or temporal lobe. Although sleep onset is in a completely different stage than NREM or REM sleep, these findings can help shape theories for dreaming and memory processing during Stage 3 of sleep. The nature and function of dreaming is a problem almost too difficult to solve, due to the lack of advanced technology. Studying different and simpler variations of dreaming can assist society in answering this question. (21, 28)

Dreaming as an Imaginative Experience

One of the main contradicting theories to the idea of memory processing and dreaming is the idea that dreaming is an imaginative experience. This theory is typically seen as an alternative to the idea that dreams involve real beliefs, memories, or real-life experience. Reality and dreaming are completely unrelated in this supposition. According to this theory, prolonged lucid dreams are also completely impossible. The main issue when it comes to this theory about imagination is that there is not one set theory with one set strength or scale of how imaginative dreams are and how "imagination" is used in each one, making it extremely difficult to prove. (28)

Nightmares

"Nightmares may be seen as a symptom of traumatization in both children and adults. Nightmares are assumed to be more frequent and more distressing among traumatized children and adolescents than among those without trauma. A total of 38 traumatized and 40 non-traumatized children and adolescents were surveyed.

Method: Nightmare-induced stress was measured with a questionnaire comprised of nine items (some of which concerned nightmare frequency and perceived intensity).

Results: Traumatized children and adolescents overall report a significantly higher number of nightmares, which they perceive in a more intense and frightening way than their control group counterparts. Traumatized children and adolescents reported an average of 9.7 nightmares per month compared to 1.7 in non-traumatized participants. The number of nightmares in traumatized girls was twice that in traumatized boys.

Conclusions: Traumatized children and adolescents report fears of dream repetitions as well as feelings of helplessness during the nightmares and are thus influenced during the daytime significantly more often. Traumatized patients have more life-threatening and violent dreams. An increased frequency of nightmare and distress is also a specific feature in traumatized children and adolescents." (17)

As portrayed in this study, traumatized patients are more likely to have nightmares than anybody else. This proves that dreaming is definitely somehow related to memory, because if the brain were "resting" or absorbing electrical impulses to create dreams, then there would be no incentive for the subconscious to bring back a traumatic memory, especially when in most cases subjects does not want to confront that trauma, especially when they are sleeping. Our brain seems to be creating some type of outlet for us humans to be able to face reality as socially and emotionally developed creatures. That is why many PTSD (post-traumatic stress disorder) patients cannot sleep very well. Their subconscious keeps trying to get them to face their trauma and overcome it by replaying it, either as a narrative or simply through flashbacks. (20)

The Senses and Dreaming

As our brain attempts to create this virtual reality about 90 minutes after we close our eyes, our whole body seems to be shutting down. (1) Our subconscious takes over and we enter a whole new world. So the question that remains is, what brings us back from this virtual reality?

Aside from the sandman switch actually chemically telling our brains to wake up (16), our senses, especially the auditory system, plays a huge role in bringing us back to the real world. (10) The auditory system is actually using the most energy of all the senses during sleep. (10) This external stimuli remains open, literally and figuratively, while the rest of one's body, including the senses, are devoting a certain amount of energy to one's sleeping state. For example, when we sleep, we use an alarm to wake up instead of a big flashing light, or when a dog sleeps and hears anything that might possibly be a threat, the dog raises one of its ears (like a reflex), while it is still sleeping. (5,10)

Discussion

Science has made it clear that every single action that is taken by the human body is for a reason. There are minimal to no excess processes that happen in our body. In simpler terms, nature makes sure that everything happens for a reason.

Both of the two main theories talked about, activation synthesis and threat simulation, have one major problem: they are only focused on one aspect of dreaming rather than dreaming as a whole.

Activation synthesis states that dreaming occurs only to absorb and make sense of excess electrical impulses.(2) This might be one small aspect of why dreaming occurs, but it is definitely not the whole picture. The brain would not create visual imagery and narratives and paralyze the muscles simply to decipher neural impulses.

The threat simulation theory seems to be more "wide-ranging" to me simply because it treats dreaming like a process rather than an effect. It provides a reason for the visual imaging and narrative aspect of dreaming, but it is not an encompassing theory either. As I stated above under "Drawbacks of the threat-simulation theory," this theory narrows the function of dreaming down to just survival-based dreams and revolves around the one emotion of fear. However, this theory doesn't account for all the bizarre dreams that occur in each person's life or why our brain chooses to remember some dreams rather than others.

In order to ensure that I came up with an all-encompassing hypothesis, I chose to use the model dream theory as a guideline.

The model dream theory has managed to compile all the different questions we have about dreaming into three main questions.

- How does neurophysiology set the stage for memory selection?
- How do dreams favor associative processes that produce bizarre and sometimes unrecognizable representations of memories?
- How do those elements combine with others in a narrative with high emotional content? (21, 22, 24)

These three questions generalize all the different possible questions we might have about dreaming and why it occurs. This paper attempted to gather information to answer each of these questions and formulate a hypothesis that would lead to more specific dream-based inquiries.

Dreaming has even been known to alter the strength of associative, procedural, and declarative memory. Even some forms of memory consolidation occur at this stage. For example, after someone experiences something unusual or has been given a heavy task load, they are known to spend more time in REM sleep than usual, most likely because dreaming is how we process our information at the end of the day. Some studies even suggest that if one reads something prior to going to sleep, they will understand it in more depth in the morning. (28) Lack of sleep, on the other hand, actually hinders the mind from correctly processing emotions and procedural learning. (15) If we take a look at those who don't have the ability to dream, such as those people with Charcot-Wilbrand syndrome, some cases have emotional and social deficits in the long run that can prevent them from having a normal life. (21)

Sleep deprivation, also a significant yet unfortunately overlooked problem, hinders procedural learning and simple processing of information. Sleep is a process that needs a certain amount of time to be completed. Stage 1 and 2 are like the "preparatory stages," Stage 3 can be seen as the "processing stage," and Stage 4 can be seen as the "refresh stage." Although Stage 4 is relatively independent from dreaming, it is the most "peaceful" stage of sleep. I can speculate that during this stage of sleep, the body could be refreshing itself and the lymphatic system could be flushing out waste and delivering nutrients because studies show that sleep deprivation causes "fogginess" in the brain due to excess waste in the brain. (1, 10, 28) Based on my own speculation, I believe deep sleep is when the cleansing happens. The lymphatic system seems only to be able to use cerebral spinal fluid to clean out the brain when we are asleep. Our wounds tend to heal faster and our hormones tend to "settle down." In short, during deep sleep our body may be at peace, but our immune system is more active. (10)

Also, severe REM sleep-deprivation is increasingly correlated to the development of mental disorders. During REM sleep, dreams help connect emotions, memories, and experiences. The parts of the brain that are active during this process only reinforce the functions I have mentioned. For example, the amygdala, which is seen as the emotional center for most humans, (20, 28) is active and related to memory processing and dreaming during REM sleep.

However, as stated before, REM sleep and dreams are not synonymous. If people have a neurological deficit and because of that cannot reach REM sleep, they can still dream in other stages of sleep; but these people tend to dream only once a week and remember significantly fewer dreams. They also tend to have major emotional and social deficits. (21) In other words, in other stages of sleep, dreams seem to be fainter and less like narratives.

I would speculate that one of the main reasons our brain seems to use a narrative format to process memories, instead of presenting us with a flash of pictures to remember, is that it is human nature to be able to remember certain events in more detail if they are showcased in narrative format. Again, since nature is efficient in everything that it does, this memory process of creating visual imagery and a story during dreaming must be the most productive way of recollection.

Theories about Emotional Processing by using dreams and memories do not always present "dreaming" as a pleasant endeavor. For example, the threat simulation theory is definitely tailored to bad dreams and nightmares. Simply because of "the high prevalence of negative emotions and threatening dream content, the threat simulation theory suggests that the evolutionary function of dreaming lies in the simulation of ancestral threats and that the rehearsal of threatening events and avoidance skills in dreams has an adaptive value by enhancing the individual's chances of survival." (28)

The emotional processing that occurs during nightmares has a lot to do with traumatic experiences. As previously mentioned, when PTSD patients sleep, their subconscious tends to present their trauma in their dreams. A lot of PTSD patients even begin to hallucinate due to the massive amount of pain their mind does not want to process. However, it is important to note that not all visual imaging is "dreaming." (12, 15, 26)

As explained before, dreaming is an evolutionary advantage and therefore any type of visual image that is created by the brain that threatens the organism's existence is not a dream. This idea is called "the vigilance hypothesis" because those types of sensations, like hallucinations, would compromise vigilance. (28) This hypothesis helps narrow down the different types of dreaming by essentially "drawing a line" and specifying that phenomena like hallucinations are not part of the dreaming process.

On a more philosophical level, life is seen as a narrative. Since EEG waves during REM sleep and wakefulness are increasingly similar, it makes perfect sense that dreams would be formatted as narratives. (19) Our brain is essentially creating this virtual reality and helping us learn, grow, and process emotions as a human being.

V. Conclusion

As humans we essentially live for our dreams, neurologically and figuratively. Ambition, drive, emotion, and the ability to use those things to achieve our goals is what differentiates us from other simpler mammals. All these different dream-related processes that help us instill knowledge into our brain and use that knowledge from an emotional perspective, is what makes us, arguably, the most powerful species on Earth. As a powerful species, we need space to safely develop our ideas and process our information. Our mind is the only space where there are no extraneous factors and we can truly dream, learn, and grow.

References

1. Association, A. S. (2016). What is sleep? Retrieved September 24, 2016, from https://www.sleepassociation.org/patients-general-public/what-is-sleep/
2. Brain activity during sleep. (2012, April 1). Retrieved September 24, 2016, from Society for Neuroscience, http://www.brainfacts.org/sensing-thinking-behaving/sleep/articles/2012/brain-activity-during-sleep/
3. Breus, M. J. (2015, May 25). How do scientists study dreams? Huffington Post. Retrieved from http://www.huffingtonpost.com/dr-michael-j-breus/dream-research_b_7306396.html
4. Campbell, I. G. (n.d.). EEG recording and analysis for sleep research. . Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2824445/
5. Desseilles, M., Dang-Vu, T., Sterpenich, V., & Schwartz, S. (2010). Cognitive and emotional processes during dreaming: A neuroimaging view. Consciousness and cognition., 20(4), 998–1008. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21075010
6. Devlin, H. (2016, May 17). What is functional magnetic resonance imaging (fMRI)? Retrieved September 24, 2016, from http://psychcentral.com/lib/what-is-functional-magnetic-resonance-imaging-fmri/
7. Domhoff, W. Finding meaning in dreams: Chapter 6. Retrieved September 24, 2016, from http://www2.ucsc.edu/dreams/Library/fmid6.html
8. Harrison, H., Hurd, R., Wargo, E., Rider, R., Christina, Ma`ruf, M. A., … Bidegain, M. (2010, January 10). Dream studies press. Retrieved September 24, 2016, from http://dreamstudies.org/2010/01/07/neuroscience-of-dreams/
9. Horikawa, T., Tamaki, M., Miyawaki, Y., Kamitani, Y., Laboratories, A. C. N., 0288, K. 619 -, … Technology, C. (2013). Neural Decoding of visual imagery during sleep. Report, 340(6132), 639–642. doi:10.1126/science.1234330
10. (J. Kanwal, personal communication, , 2016)
11. Kussé, C., Muto, V., Mascetti, L., Matarazzo, L., Foret, A., Bourdiec, A., & Maquet, P. (2010). Neuroimaging of dreaming: State of the art and limitations. International review of neurobiology., 92, 87–99. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20870064
12. Lippman, A. (1994). The Neurophysiology of sleep and dreams. Retrieved September 24, 2016, from Bryn Mawr, http://serendip.brynmawr.edu/bb/neuro/neuro03/web1/alippman.html
13. Mandal, A. (2015, October 27). Dopamine functions. Retrieved September 24, 2016, from Health, http://www.news-medical.net/health/Dopamine-Functions.aspx
14. Marzano, C., Ferrara, M., Mauro, F., Moroni, F., Gorgoni, M., Tempesta, D., … De Gennaro, L. (2011). Recalling and forgetting dreams: Theta and Alpha Oscillations during sleep predict subsequent dream recall. The Journal of Neuroscience, 31(18), 6674–6683. doi:10.1523/JNEUROSCI.0412-11.2011
15. Mastin, L. (2013). Sleep - types and stages of sleep - REM sleep. Retrieved September 24, 2016, from http://www.howsleepworks.com/types_rem.html
16. News, N. (2016, August 3). Researchers discover Sandman's role in sleep control. Retrieved September 11, 2016, from http://neurosciencenews.com/sandman-sleep-control-4774/In-line Citation:(News, 2016)
17. Ossa, F., Bering, R., & Pietrowsky, R. (2013). [Prevalence and intensity of nightmares in traumatized versus non-traumatized children and adolescents]. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie., 41(5), 309–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23988833

18. Parmeggian, P. (2016). Systemic Homeostasis and Poikilostasis in sleep. Retrieved September 24, 2016, from http://www.worldscientific.com/worldscibooks/10.1142/p720
19. Shank, S., & Margoliash, D. (2008). Sleep and sensorimotor integration during early vocal learning in a songbird. Nature., 458(7234), 73–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19079238In-line Citation:(Shank & Margoliash, 2008)
20. Society, S. R. SLEEP - thematic and content analysis of Idiopathic nightmares and bad dreams. Retrieved September 24, 2016, from http://www.journalsleep.org/ViewAbstract.aspx?pid=29326
21. Stickgold, R., Hobson, J. A., & Fosse, R. (1972). Sleep, learning, and dreams: Off-line memory Reprocessing. Science J. Exp. Psychol. Behav. Brain Res. J. Cognit. Neurosci. Nature Neuro-sci. Electroencephalogr. Clin. Neu-rophysiol. Science Brain Res, 96(130), 321–302. Retrieved from http://www.cogsci.ucsd.edu/~pineda/COGS175/readings/Stickgold.pdf
22. The brain as a dream state generator: An activation-synthesis hypothesis of the dream process (1977). American Journal of Psychiatry, 134(12), 1335–1348. doi:10.1176/ajp.134.12.1335
23. Valli, K., Revonsuo, A., Pälkäs, O., Ismail, K., Ali, K., & Punamäki, R. (2005). The threat simulation theory of the evolutionary function of dreaming: Evidence from dreams of traumatized children. Consciousness and cognition., 14(1), 188–218. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15766897
24. van der Linden, S. (2016). The science behind dreaming. Retrieved August 17, 2016, from http://www.scientificamerican.com/article/the-science-behind-dreaming/
25. Vinogradov, A. E. (2003). DNA helix: The importance of being GC-rich. , 31(7), . Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC152811/
26. Walker, M. P. (2008). Sleep-dependent memory processing. Harvard Review of Psychiatry, 16(5), 287–298. doi:10.1080/10673220802432517
27. Whitmann, L., S., & Ciric, P. (n.d.). Typical dreams: Stability and gender differences.
28. Windt, J. M. (2015, April 9). Dreams and dreaming. Retrieved September 24, 2016, from http://plato.stanford.edu/entries/dreams-dreaming/

Visual Aid References

29. Eakin, E., Ryback, T. W., Scott, A. K., Heij, P., Tolentino, J., Angell, R., … Osnos, E. (2015, November 25). Emily Eakin. . Retrieved from http://www.newyorker.com/magazine/2015/12/07/bacteria-on-the-brain
30. Normal sleep EEG: Overview, stage I sleep, stage II sleep. (2016, August 24). Retrieved September 25, 2016, from http://emedicine.medscape.com/article/1140322-overview
31. TheLucidDreamSite. (2013). The Dorso-Lateral Prefrontal-Cortex and lucid dreaming. Retrieved September 25, 2016, from http://theluciddreamsite.com/the-dorso-lateral-prefrontal-cortex-and-lucid-dreaming.html
32. Sleep Aid Sounds. (2012). The Sleep Cycle. Retrieved December 26, 2016, from http://sleepaidsounds.com

A World of Possibility: Tier-Oriented Base-Storage Network for CCN Routing

Yuchen Xu

Author background: Yuchen Xu grew up in China and currently attends The High School Affiliated to Renmin University of China, located in Beijing, China. His Pioneer seminar topic was in the field of computer science and titled "Rethinking the Internet Architecture."

Abstract

Since Jacobson et al.'s paper (2009), Content-Centric Network (CCN) has become one of the most popular new network approaches, and routing in CCN has attracted attention from researchers all over the world. Proposals for possible routing mechanisms include flooding, tag-based routing and name-based routing using distance. In this paper, we briefly analyze the three existing models, and propose a Tier-Oriented Base-Storage (TOBS) network. In TOBS, routers are strictly organized into tiers, and base routers store content apart from caching. We give a detailed explanation of the TOBS model, indicating the benefits from a clear organization. We also relate TOBS to other models and point out the limitations of TOBS for future improvements.

Keywords

Content Centric Network; Routing

I. Introduction

1.1 Content Centric Network

The concept of the "Information Centric Network," which is expected to enhance network efficiency, has been researched and discussed for a few years. One of the proposals that put this network into reality, or namely, allowed for potentially more efficient networking, is the Content Centric Network (CCN). In "Networking Named Content" (Jacobson 2009), researchers provided a solution by establishing a node model and analyzing the transportation, routing and security of the new network, proving the usefulness of their model. The model will be described in more detail in the second section of this paper.

1.2 Routing Problem

Despite the wide-scale coverage of that pioneering paper, many issues and mechanisms within this new network remain unclear or unsolved. Among these is routing, the mechanism needed to direct certain packets in desirable ways. Jacobson's paper (2009) gave a vague description of the mechanism: "Any routing scheme that works well for IP should also work well for CCN." Instead, it puts more focus on the transformation from current networks to CCN, stating that optimal topology can be constructed dynamically in CCN.

1.3 Routing Proposals

But what about the ways to route content? Researchers have different ideas. The foci of this research mainly fall on content headers and routing topology.

Although the prefix system in IP can be readily used in CCN, there are also other kinds of headers, such as tags (Papalini et al., 2014), which give a new possibility in routing identifiers. Ideas of routing topology include flooding (Wang et al., 2015), the most basic method of routing content. These existing models will be discussed in the third section of this paper.

1.4 Tier-Oriented Base-Storage network

In an effort to shed light on how to tackle the routing problem, we propose a different model—the Tier-Oriented Base-Storage network—in the fourth part of this paper. In this network, routers and content objects are both identified by ordered sequences (like numbers, as is the case in the illustration), and we organize routers within the network into tiers; each tier signifies different levels of accuracy when identifying content objects, so that the network is highly organized. We also make use of Tier-1 routers (at the bottom of the network) for solid storage of content (contrast to temporary cache), so that all content can be found regardless of popularity, and new content can be placed on file in a deterministic way. To reduce the amount of information to be transmitted, we establish a "parent-daughter" relationship between routers on different tiers, so that parents can make routing decisions without knowing the specific content in lower-tier routers. We also give an analysis of the TOBS model.

1.5 Related Work

To broaden our view, we relate our model to a few network approaches outside of CCN routing in the fifth part of the paper, including Chord, Distributed Hash Table and Wireless Ad Hoc Network. These models are essentially different from CCN mechanisms, but the ideas in them may provide insight into future CCN routing research.

II. Background

With the rapid development of technology, the Internet has been increasingly accessible to common computer users, and it has been one of the major ways of distributing content. Currently, the network is host-oriented, which means communication of information is based on the connection between hosts. However, given all kinds of delays within the network and the fact that delays are positively correlated to the physical distance and the number of routers between hosts, the efficiency of content distribution has become one of the main issues in the current network, with the number of hosts growing at an enormous rate. Other problems also emerge, such as the availability of new IP addresses to identify different hosts. A different approach to networking is needed.

Jacobson (2009) proposed a solution to the CCN network in "Networking Named Content." The paper covers a wide range of topics, picturing different sides of the CCN statue, part of which is described below.

2.1 What is CCN and why CCN?

"People value the Internet for *what* content it contains, but communication is still in terms of *where*." So "the direct, unified way to solve these problems is to replace *where* with *what*." (Jacobson 2009)

This statement accurately describes the difference between our current network and CCN. Our current network is based on establishing connections between nodes (for example, connecting your computer to http://www.ccnx.org) before exchanging information packets (such as a specific research paper), while CCN mainly focuses on finding desired information packets (the research paper) in the network.

In short, “CCN talks about data, not to nodes.” (Jacobson 2009)

2.2 CCN Packets

Interest packets include content name and selectors; data packets include content name, data and other identifying information.

2.3 Routers in CCN

Routers have three main parts:

The Forwarding Information Base (FIB), by name, forwards interest packets to possible sources.

The Pending Interest Table (PIT) keeps track of messages forwarded and where they come from, in order for the router to get the data back to the original router that wants the content. These two parts construct the very basics of routing in CCN.

The Content Store (CS) serves to store content objects (Jacobson 2009), a caching function similar to buffer memory, which increases efficiency for the entire network. Caching is yet another popular issue in CCN research, which includes where to cache what content objects according to what standard, etc. Caching is included in the TOBS model to be described later, but will not be the focus of this paper.

III. Analyses of Different Routing Methods

3.1 Scoped-flooding model

3.1.1 Flooding

Flooding is one of the simplest approaches to routing in network. Normally, flooding means spreading packets to adjacent routers indiscriminately. For all the ease it brings to network protocols, network throughput increases exponentially as the size of the network grows. Thus, flooding on a large scale is obviously unrealistic, at least under current technologies.

3.1.2 Model Description

However, flooding is still worth considering, and in “Pro-Diluvian: Understanding Scoped-Flooding for Content Discovery in Information-Centric Networking,” (Wang 2009) researchers studied flooding on a small scale. By excluding clients, they established a node-centric ring-based model, which is basically “circles” of nodes around the central node. Under this model, they are able to calculate the neighborhood growth rate and show that the rate can be estimated using local information close to the node. Researchers further analyzed flooding strategies and pointed out that dynamic flooding with a specific radius derived at each node and position taken into account is effective in heterogeneous topological structure (where topological characters vary from place to place). They concluded that flooding, when restricted at an optimal radius (quite small) they derived, could allow the network to have the most gains.

3.1.3 Limitations and Prospects

That flooding works well at network edges may be true, but taking into account the costs required for small-scope flooding to be adapted into the network, it remains a problem whether this flooding model is applicable and economic. Still, the analysis of flooding can possibly promote later study of a better defined “flooding” strategy that can benefit the network on a large scale.

3.2 Tag-based multiple-tree routing model

In "Scalable Routing for Tag-Based Information-Centric Networking,"(Papalini 2014) researchers proposed a model that basically makes use of multiple trees to route packets.

3.2.1 Tags

Multiple trees are nothing new, but the "tag" component in their model is worth noting. Instead of using IP prefixes, researchers defined "content descriptors," which are virtually a set of string tags. These tags are more expressive due to flexibility—a match of tags simply requires affiliation, while a match of prefixes requires order of components, additionally. To give an analogy, tag matches are like determining the relationship between a set and its subset (which are both characterized by randomness, namely {1,2} and {2,1} are the same), while prefix matches are like determining the relationship between two sequences (which takes permutation into account). In this way, a wider range of descriptors can be included in a single set of tags. (Content identifiers and locaters are also parts of the packet header.)

3.2.2 Scalability

The general idea of this model is to "use descriptors to find an object," which mainly includes directing the request to content providers and getting the content back to the user. In the demonstration, a tree of routers with their FIBs indicating the direction of other routers is displayed. It is doubtful whether similar FIBs still work under a large-scale network; despite the fact that the researchers mainly analyzed small-range models in the paper, they proved the theoretical feasibility (scalability) by showing the limited growth of required memory with the growth of users to hundreds of millions.

3.2.3 Request/Reply and Subscribe/Publish

There is another feature worth noting in this model: it includes both a request/reply model and a subscribe/publish model, namely "pull" and "push." Both work in similar ways, but the idea the model has in it can potentially be useful for optimization of CCN—in Jacobson's original paper, CCN is a network mostly oriented for obtaining information (via interest), and not intended for the content providers to send content back to users, a drawback that stops CCN from completely replacing current networks. This model may shed light on future research on related issues.

3.3 Distance-based Content Routing (DCR) model

3.3.1 Model Description

As suggested by name, DCR makes use of distance between routers for routing. Each network node within the network, as well as each Named Data Object (NDO), is assigned a name, which can either be flat or hierarchical. Apart from opportunistic caching in all routers, content objects are stored at specific nodes, named "anchor." Anchors update information to other routers for them to decide which path to route requests in the network. (Garcia-Luna-Aceves, 2014)

3.3.2 Comments

Researchers admit the lack of scalability analysis in the conclusion of their paper—increased number of data objects, and increased network size, which can lead to an enormous throughput for update messages, all pose threat to the network model.

However, we should still notice the positive sides of this model—the idea of storing content at specific anchors, and giving routers names contribute to the development of our routing model, to be discussed in the next section.

IV. A New Outlook—Tier-Oriented Base-Storage Network (TOBS)

In this section, we introduce a new model of routing in CCN—Tier-oriented Base-storage Network (TOBS). We will first give a brief overview of the model, then analyze its working mechanisms within a simplified model, and view the model on a larger scale.

4.1 Overview

4.1.1 Symbols

The TOBS model is mainly concerned with the organization and usage of routers within the network. To better describe the model, we assume that each piece of content object has a serial number unique to it (represented by C123, C2098 etc., where C is for Content). We also assume that each router has a serial number ("name") unique to it (represented by R1, R11, R12 etc., where R is for router).

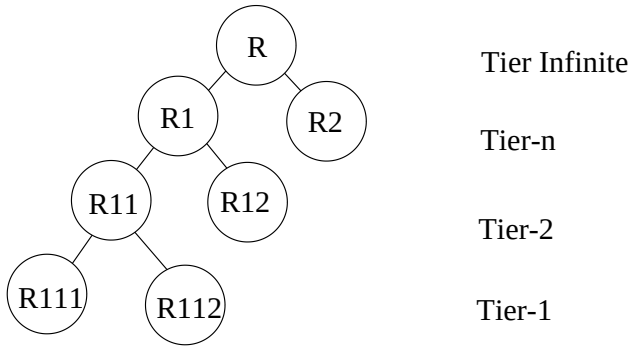


Figure 4-1

4.1.2 Tier Orientation

Routers are organized into different tiers according to their names (see figure 4-1; router R is special because it does not have a specific name). Routers on the same tier are called "neighbors." (R11 and R12 are neighbors) Upper tier routers are "parents" of lower tier routers and lower tier routers are "daughters" of upper tier routers (R1 is the parent of R11 and R11 is the daughter of R1).

Neighbors on Tier 1 are interconnected, so that one router can get to its neighbor without leaving the tier, but the router is not necessarily connected to all neighbors (see figure 4-2; both orientations satisfy the requirements of this model).

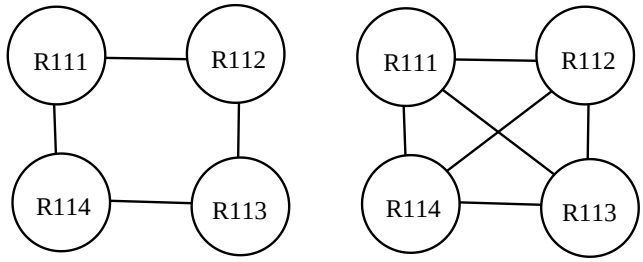


Figure 4-2

4.1.3 Model Routers Within the Network

Figure 4-3-1 depicts a typical router on Tier 1. The router is connected to its neighbor(s), as is described earlier; it is also connected to its parent. Within the router, the caching function in Jacobson's paper is preserved; and the information table serves to keep track of the name of connected routers and the basic characteristics of content objects within the router (which we will discuss in part 2).

"Base storage" is one of the defining characteristics of the TOBS model. All routers on Tier 1 store content objects according to the "longest match" between router names and content object serial numbers. Thus, in Figure A, all objects starting with C111 (including C1112, C111939, C111, etc., but not C110 or C1101) will be stored at R111.

Figure 4-3-2 depicts a typical router on upper tiers. The router is connected to its daughters and parent. The caching function is also preserved here; the information table keeps track of the names of connected routers (especially the names of daughters) and basic characteristics of daughters.

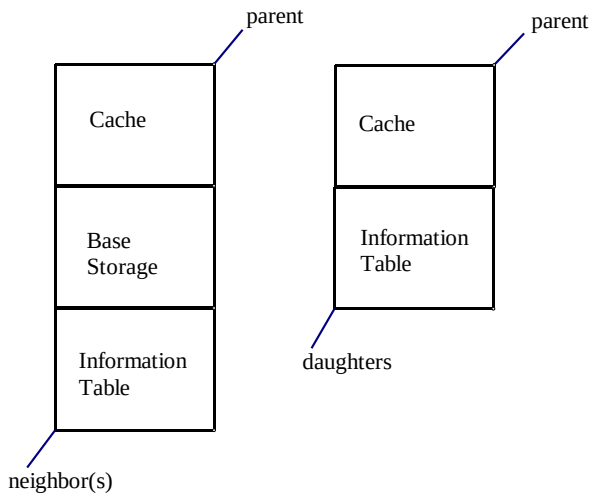


Figure 4-3

4.2 Simplified Model: Seeing the forest through the trees

With the basic concepts in part 1 in mind, we will go on to discuss how these concepts allow the TOBS model to work as a network. We start with basic routing function of the network, and then discuss how the network reacts to changes within it.

Figure 4-4 shows a simplified version of a TOBS network. We will use this figure for the first to third part of the discussion.

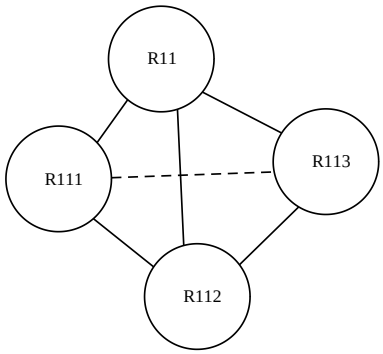


Figure 4-4

4.2.1 Requesting Information

The request for information is one of the most important parts of CCN. In TOBS, every router can send out requests for specific content (with the serial number of the wanted content object).

If the starting router is a Tier-1 router, it first checks whether the object number matches its own name—if there is a match, then the router can just look for the content within its own storage. Otherwise, the request is forwarded to its parent to check for a match, and if there is a match, the parent router forwards the request downward to the matching daughter; if not, then the request is further forwarded upward, until there is a match, or the request comes to Tier Infinite (when Tier Infinite helps the request find a match in its daughters).

4.2.2 Adding Information

Rather than spreading content, the initial CCN proposal mainly focuses on looking for content, because the convenience of looking for content is where CCN outweighs the current network. However, in TOBS, the clear organization of routers makes it easy to add new content to the network. When new content is added from any node within the network, the content object can follow a similar route of a request to find where the content should be stored.

4.2.3 Addition and Deduction of Routers (Figure 4-5)

Under the status quo, the number of routers is growing at an extremely high rate, so it is important for a network to be highly kinetic. In the following subsections, we discuss small- and large- scale addition and reduction of routers.

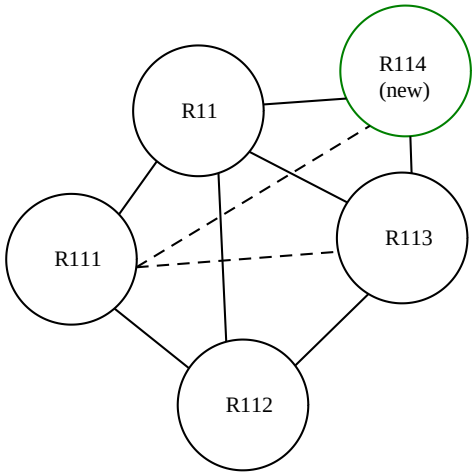


Figure 4-5

4.2.3.1 Addition

We consider the situation of adding one router to the TOBS network. Usually, the new router is added to Tier-1, and the specific location of addition can be decided by physical adjacency. The new router possesses a similar function to existing Tier-1 routers.

The newcomer first informs its parent of its presence, and then the parent will decide on one other daughter to share storage with the newcomer. The way that the parent decides the specific content to be transmitted from an existing daughter to the newcomer is by measuring the total popularity of content in each daughter (a piece of information that other daughters have updated the parent about) and to allot some "popularity" to the newcomer so as to reduce the number of requests aimed at one specific daughter.

4.2.3.2 Further Addition

If the number of content objects and the number of routers grow too sharply, a "new Tier-1" under "existing Tier-1" may be needed. However, this change can be a huge shift—not only will all the content objects be transmitted, but also the connection will undergo sharp change (because Tier-1 routers are interconnected while routers in other tiers are not, so all "existing Tier-1" neighbors will need to disconnect and new connections need to be established, which can be costly).

This is not a big concern in TOBS, because storage costs are relatively low in today's network, so new content can still be stored at "existing Tier-1." However, an increased number of objects at one router can mean increased overhead, whose cost cannot be ignored. Higher-efficiency transmission is a potential solution to this problem.

4.2.3.3 Reduction

Sometimes, routers in TOBS get turned off manually. Before people do so, however, measures need to be taken to avoid hurting the whole network.

If a Tier-1 router is taken away, the router needs to inform its parent first to use a similar mechanism to redistribute the router's storage to other daughters.

If a higher tier router is taken away, the best solution is to replace the router with a similar one—this seems to be nonsense. There are other solutions, which we will talk about in the following subsection.

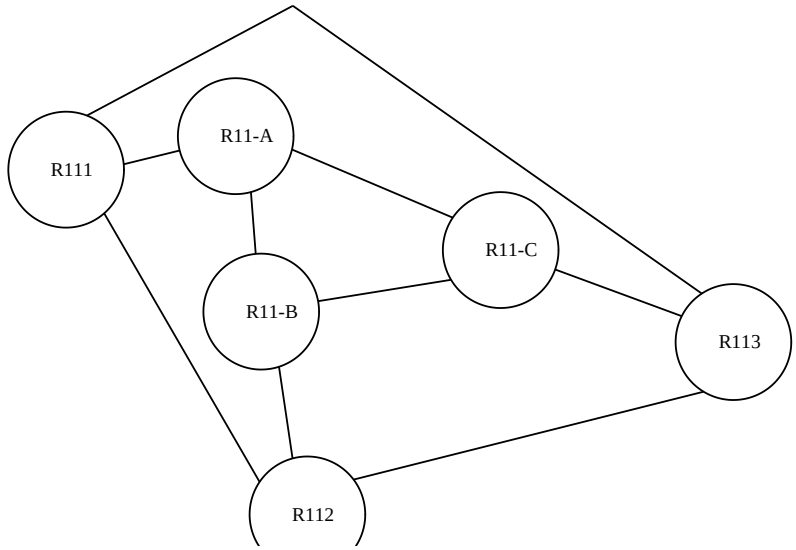


Figure 4-6

Figure 4-6 shows an enhanced version of the TOBS network. We use this figure for the last part of the discussion.

In figure 4-5, parents are assumed to be "single" routers—namely, there is only one router in the network bearing the name R11. This is a simple assumption, but if we take into account the limited number of connections a router can have with other routers, one parent may not be enough for a group of daughters.

To solve this problem, we introduce multiple routers to serve the role of a single parent (each router is called a "sub-parent," represented by R11-A, R11-B and R11-C). Each sub-parent is connected to a few daughters, and the sub-parents are interconnected.

4.2.4 Single Node Breakdown

As is common to most networks, TOBS needs to face the problem of node breakdown within the network. We will discuss the case of single node breakdown here, because the chance of multiple nodes serving a similar function breaking down at the same time is negligible.

4.2.4.1 Detecting Breakdown

When the message from one router fails to reach the desired destination, it is highly likely that a breakdown has occurred either within the destination router or on the route between the two routers. The starting router can notice this via different mechanisms, including timeout.

4.2.4.2 Dealing With Breakdown

After a breakdown is detected, a message is sent to the Internet Service Provider (ISP, which can be Tier Infinite in this case). The message has an empty serial number (which leads it up to Tier Infinite), as well as the name of the router that broke down, so that the ISP can work on fixing the malfunctioning router.

It is helpful to point out two important points here. Firstly, because there are multiple sub-parents working under the same router name on each tier other than Tier-1 and the chance

that sub-parents break down at the same time is relatively low, the event in which the connection between different tiers is completely blocked is unlikely. Secondly, when a daughter fails to send the "breakdown" message directly to the parent with which it is connected, the daughter can flood the message to its peers (neighbors on Tier-1 or other sub-parents on other tiers) until the message goes up to a functional parent.

4.3 Whole Picture: Scaling up

We have talked a lot about the basic mechanisms within the TOBS network, most of which are based on simplified models. However, as we discussed in part III of this paper, scalability is one of the most important issues when considering routing models in CCN. Thus, we will scale up and discuss the advantages and disadvantages of the TOBS network on a larger scale.

4.3.1 Setting up the Network

In reality, it is preferable to distribute all routers on and below Tier Infinite to network users. Although this method can cause some users to be with routers storing content and others to be with routers only with cache, the method still prevents ISPs from holding too many routers. High-frequency users can possess Tier-1 routers, and low-frequency users can have upper tier routers.

4.3.2 Tier Ultimate

Tier Ultimate stands for the tier that lies above Tier Infinite. Due to physical distance, it is unrealistic to have all content objects under one series of routers (like the ones shown in figure 4-1). Instead, we put different series of routers in different physical regions, and connect their Tier Infinite to Tier Ultimate. When Tier Infinite fails to find the requested content within one series of routers, it forwards the request to Tier Ultimate, and Tier Ultimate further forwards the request to other Tier Infinite routers (by flooding or other methods).

4.3.3 Analysis

The TOBS network has two significant advantages:

Firstly, upper tier routers can route request without knowing the specific serial numbers of objects stored in Tier-1 routers, because routing is mainly based on longest match.

Secondly, the network has clear organization, which means it is highly deterministic. This trait prevents the extra cost brought about by opportunistic networks, and allows rare content to be found accurately.

However, there are also drawbacks that await further improvements.

Firstly, the routing cost cannot be ignored under the current situation. If there is a relatively big number of tiers, request and content objects may need to travel a long distance before reaching the destination. And the physical distribution is also a limiting factor when establishing the network. Both problems increase the routing cost.

Besides, Tier Ultimate is not strong enough as a component in the network. With increased communication between different regions across the world, requests for information from a different region will increase significantly, and Tier Ultimate needs further improvement to support these communications.

V. Related Work

5.1 Chord

The TOBS model is, in some ways, similar to peer-to-peer network. As "a fundamental problem that confronts peer-to-peer applications is to efficiently locate the node that stores a particular data item," the model of Chord provides a solution: "given a key, it maps the key onto a node." (Stoica et al., 2003) Chord displays an evenly distributed structure (as opposed to the highly organized structure in TOBS), theoretical scalability and versatility under change.

5.2 Distributed Hash Table

Distributed Hash Table (DHT) is one of the routing topologies in peer-to-peer (P2P) networks. Due to the highly dynamic nature of P2P networks, any protocol within them should address this problem effectively, as what researchers have done in Chord. Manku's paper (2003) provides a solution by using adapting classical parallel interconnection networks.

5.3 Wireless Ad Hoc Network

Totally contrary to the deterministic structure in TOBS, wireless ad hoc network is highly opportunistic. It takes into account dynamics of topology within the network.

The most innovative part in this network comes with the elimination of routers—packets travel through devices directly. Since the radius of network is limited, when a destination is outside of network, the network stores the content object it wishes to spread at the node that is most likely to travel to the desirable destination (which is analogous to the spread of epidemic). (Johnson& Maltz, 1996)

VI. Conclusions and Future Work

We briefly reviewed the CCN proposal with focus on routing mechanisms, and analyzed some existing models including tag-based routing and scoped flooding. We also explained in great detail our proposal of the Tier-Oriented Base-Storage (TOBS) network, which has the advantage of accuracy and all-coverage of looking for content and saves throughput by not letting parents know the exact content objects within daughters. We realize that there are still limitations to the TOBS model, and would like to direct future work to the following areas:

1) Using quantitative models to further evaluate the applicability of theTOBS model
2) Lowering or controlling the cost of message travelling through a long path
3) Solving the problem of Tier Ultimate in TOBS and replacing it with a more efficient transmitter

References

Jacobson, V., Smetters, D. K., Thornton, J. D., Plass, M., Briggs, N., &Braynard, R. (2012). Networking named content. *ACM Conference on Emerging NETWORKING Experiments and Technology, CONEXT 2009, Rome, Italy, December* (Vol.55, pp.117-124).

Wang, L., Bayhan, S., Ott, J. &Kangasharju, J., Sathiaseelan, A., &Crowcroft, J. (2015). Pro-Diluvian: Understanding Scoped-Flooding for Content Discovery in Information-Centric Networking. *International Conference on Information-Centric NETWORKING* (pp.9-18). ACM.

Garcia-Luna-Aceves, J. J. (2014).Name-based content routing in information centric networks using distance information. *The, International Conference* (pp.7-16).

Papalini, M., Carzaniga, A., Khazaei, K., & Wolf, A. L. (2014).Scalable routing for tag-based information-centric networking. *The, International Conference* (Vol.97, pp.2157-2166).

Stoica, Ion, Morris, Robert, Karger, David, Kaashoek, M. Frans, &Balakrishnan, Hari. (2003). Chord: a scalable peer-to-peer lookup service for internet applications. *IEEE/ACM Transactions on Networking, 11*(4), 149-160.

Manku, G. S. (2003). Routing networks for distributed hash tables. *PODC '03 Proceedings of the twenty-second annual symposium on Principles of distributed computing*, 133-142.

Johnson, D. B., &Maltz, D. A. (1996). Dynamic Source Routing in Ad Hoc Wireless Networks. *Mobile Computing* (Vol.353, pp.153-181).

Vishnumurthy, V., & Francis, P. (2007). A comparison of structured and unstructured P2P approaches to heterogeneous random peer selection. *Usenix Technical Conference, June 17-22, 2007, Santa Clara, Ca, USA.*

Comparing and Contrasting Economics Development: the Case of Malaysia and Singapore

Alesha Wong Yun Ying

Author background: Alesha Wong Yun Ying grew up in Malaysia and currently attends Tenby International School, located in Setia Eco Park, Malaysia. Her Pioneer seminar topic was in the field of economics and titled "An Overview of the U.S. Macro-Economy."

Abstract

This paper examines the economic development of Malaysia and Singapore following their separation in 1965. It describes the chronological behaviours of the main economic indicators of economic well-being in search of plausible causes for the diverging performances of the two countries. Finally the paper uses insights from the Solow Growth Model to explore determinants of both countries' standard of living.

I. Introduction

Prior to 1965, Malaysia and Singapore were a united country. The Economist (2015) reported:

> *"When Singapore started life as an independent, separate country in 1965, Singapore's prospects did not look good. Tiny and underdeveloped, it had no natural resources and a population of relatively recent immigrants with little shared history."*

The crucial question is: how did Singapore move from being a lesser developed country with so few prospects for future economic growth to becoming one of the wealthiest nations in such a short time?

If you walked through the heart of Kuala Lumpur, the capital of Malaysia, today, you would find yourself in a vibrant city center surrounded by street stalls, shopping malls, and skyscrapers. You would encounter crowds of tourists and locals alike. One of the factors you would notice would be the variety of different ethnicities of its people such as Malays (50%), Chinese (23%), and South Indians (7%), all of which combined make up the majority of the population (Malaysia Demographics Profile, 2015). This is undoubtedly part of Malaysia's attraction as its people take great pride in living in a harmonious multi-ethnic community.

Singapore, while having a different composition in its population, similarly takes great pride in its multi-ethnic society composed of 74% Chinese, 13% Malays, and 9% ethnic Indians (Statistics Singapore, 2015). Its history is rooted in the intertwinement of the different cultures, and therefore is very similar to that of Malaysia. Not surprisingly, local food comes in a large variety of cuisines. There are other similarities about the two bustling

capitals: the atmosphere in general, the architecture, the selection of stores, or the type of food available. Hence it would be hard to analyze the difference between the two countries by just looking at Kuala Lumpur and Singapore. Instead, you have to consider Malaysia as a whole to uncover the true differences, economic and otherwise.

Think about some of the basic underlying facts regarding the two countries. Malaysia has a land area of close to 128,000 square miles and is therefore over four hundred times larger than Singapore, which is less than 280 square miles. Malaysia is home to close to 30.3 million people while Singapore's population is only roughly 5.5 million people (World Bank, 2016). Both territories have significantly different GDP per capita, with Singapore's listed at close to $52,000USD per capita (see Table 1). In contrast, Malaysia's is only $11,000 USD per capita (World Bank, 2016). Note that GDP has been converted to USD. To put matters into perspective, Singapore seems to be at the same level of development as Luxembourg, while Malaysia is more similar to Greece (International Monetary Fund, 2015). For an outsider, this outcome is unexpected, considering not only the size and demographics of Malaysia, but also the fact that it has plenty of natural resources such as tin, rubber, petroleum. Meanwhile, Singapore has very limited natural resources.

Table 1: GDP, Population, and per capita GDP, Malaysia and Singapore, 1965 and 2015, US $

	Malaysia			Singapore		
	GDP	Population	p.c. GDP	GDP	Population	p.c. GDP
1965	$2.956	9.6	$308.92	$0.974	1.9	$516.29
2015	$296.217	30.3	$9,766.17	$292.739	5.5	$52,888.72

Source*: *World Bank, GDP current US$, in Billions; Population, in Millions

The first lesson to learn from these facts is that the Wealth of Nations, as Adam Smith referred to per capita income, is not determined by population density or the amount of natural resources available to a country. While it is tempting to blame high population density on India's level of development, Singapore (and Hong Kong) are clear counterexamples. Also note that Hong Kong, like Singapore, does not have many natural resources: the SAR (Special Administrative Regions) of China does not have sufficient water for its population and always had to import water from the mainland. As a matter of fact, there is a significant amount of literature that has focused on the so-called "resource curse."

Size does not matter either. The GDP of Singapore is by now almost as large as that of Malaysia. However, there are only 5.5 million people who produce it, while for Malaysia there are 30.3 million people.[44]

The purpose of this paper is to compare and contrast the economic growth of Malaysia and Singapore since 1965, when both nations decided to part ways to become two separate countries. The paper proceeds as follows: first, I will give a brief history of the formation of the Federation of Malaya and how it fell apart. In the process I will focus on

[44] Strictly speaking, it is the employed labor force, of course, that is producing GDP, not the population.

key players and their role in the economic growth of both countries. Next, I will analyze the time series behavior of the four main economic indicators of well-being in both countries: real GDP growth, inflation, unemployment, and the exchange rate. Part of the analysis explores political and economical reasons for their behavior. Next, I will use insights from the growth literature to explain Singapore's relatively high growth rate. A final section provides concluding remarks summarizing the findings and exploring future avenues of research.

II. Brief History

Before September 1963, the Federation of Malaya (currently known as Peninsular Malaysia) was an independent nation whilst Singapore, North Borneo (currently known as Sabah), and Sarawak were British colonies. The Federation of Malaya gained its independence from Britain on the 31st August, 1957. It joined the Commonwealth of Nations (a voluntary association of 52 independent and sovereign states, most of which were former British colonies or dependencies of these colonies.) with Tunku Abdul Rahman as its first prime minister. Singapore, Sabah and Sarawak gained their independence from Britain by joining the Federation by Malaysia to form the Federation of Malaysia in September 1963. Singapore won full internal self-government from Britain for all matters except defence and foreign affairs in 1959. The People's Action Party (PAP) won a landslide general election in 1959 with Lee Kuan Yew as the Prime Minister. Later in 1965, Singapore separated from Malaysia and became an independent nation with Lee Kuan Yew of PAP becoming the first Prime Minister of independent Singapore.

Singapore's first task as an independent nation was to instill national unity and loyalty, which was challenging due to its multi-ethnic population. While quickly deciding on its new national flag, anthem and crest, it also made Malay, Tamil, Chinese, and English the official languages. However, it designated Malay as the national language because it envisioned a future merger with Malaysia.

Singapore's second, but also harder task, was to transform Singapore from an entrepôt economy dependent on the commodity trade from Peninsular Malaysia with no tradition in manufacturing to an industrialized society. The PAP continued to believe that Singapore's survival as a country depended ultimately on a merger with Malaya. Still, PAP's Minister of Finance, Goh Keng Swee, drew up a four-year development plan that included investing incentives, such as a low taxation rate, tax holidays, and temporary tariff restriction. In 1960, he stated

> *"Major changes in our economy are only possible if Singapore and the Federation are integrated as one economy. Nobody in his senses believes that Singapore alone, in isolation, can be independent."* (Country Studies, 2003).

At the same time, politicians within the United Malays National Organisation (UMNO) government led by Malaya's Prime Minister Tunku Abdul Rahman were increasingly hesitant about a future union with Singapore as they believed the PAP was extremely left wing. In April 1961, Tunku was forced to reconsider the future merger because Lee Kuan Yew's government was in danger of being brought down, and there was the distinct possibility of his party being replaced by a pro-communist government. To avoid this transition, on May 27, 1961, Tunku Abdul Rahman presented a proposal for the union of the Federation of Malaya, Borneo, Sarawak, and Singapore to become one country, called Malaysia. The Malaysia Agreement was signed by leaders of all four participating states and the United Kingdom on July 9th, 1963 (The Commonwealth, 2013).

From the very beginning, Malaysia and Singapore were fully aware of their differences in ideologies, both politically and economically. Hence, it came as no surprise to see the increasing conflicts between the two entities (Lee, 1963; 5). At the political level, the two major parties of Malaysia, PAP and UMNO, were accusing each other of "communalism": the PAP was seen as advocating the cause of the Chinese, while UMNO seemed to prefer the Malays. Racial riots soon erupted, and despite agreeing initially on a two year truce in 1964, the rivalry soon flared up again. UMNO always favoured Malays over the Chinese and Indians because the party believed that Malays were the original people inhabiting the country. This communal notion did not sit well with Singapore's leader who advocated social equality, or in other words, a 'Malaysian Malaysia' (Cheah, 2002).

In terms of economics, things were not looking all that good for Singapore as progress towards a common market (targeted to be completed in twelve months) was slow, and Kuala Lumpur was demanding that Singapore funded the common defense system. This would have required Singapore to commit the majority of its revenue towards that end. Furthermore, there was the threat to shut down the Bank of China branch, which happened to be an important financial channel between Singapore and China (Country Studies, 2003).

By the second half of 1965, the political tensions between the two countries showed no signs of getting resolved. At the Commonwealth Prime Minister's conference held in June 1965, the Malaysian Prime Minister, Tunku Abdul Rahman, decided that severing relations with Singapore was the only reasonable course of action for both countries to insure progress (Fong, 1990). Singapore and Malaysia started to draft the paperwork for their separation in secrecy. Only a few senior members of both the UMNO and PAP were involved. After long discussions between senior politicians from both sides, the final version of the separation agreement was completed on August 6th and was subsequently signed by all parties involved (The Straits Times, 1998).

On 9 August 1965, Tunku Abdul Rahman made the separation official: he announced the Constitution of Malaysia (Singapore Amendment) Bill, 1965, that would permit Singapore to leave Malaysia, and to turn itself into an independent and sovereign state. In Singapore, the announcement was aired on the radio at 10:00 am (Turnbull, 2009) and a press conference, called by the prime minister, was held at 4:30 pm that afternoon (Lee, 1998). Overwhelmed by emotion and tears, Lee Kuan Yew communicated the hopes he had about the merger, but also explained why the separation was inevitable for the country (Abisheganadan, 1965).

III. Economic Data Analysis

Having given some background regarding the origins of the two countries, this paper will now look at the historical development of four economic indicators of well-being. These are real GDP growth, the inflation rate, the unemployment rate, and the exchange rate. The analysis of these four indicators, along with an understanding of economic policies in Malaysia and Singapore, is intended to provide a better initial understanding of how the two economies have progressed since the political separation in 1965.

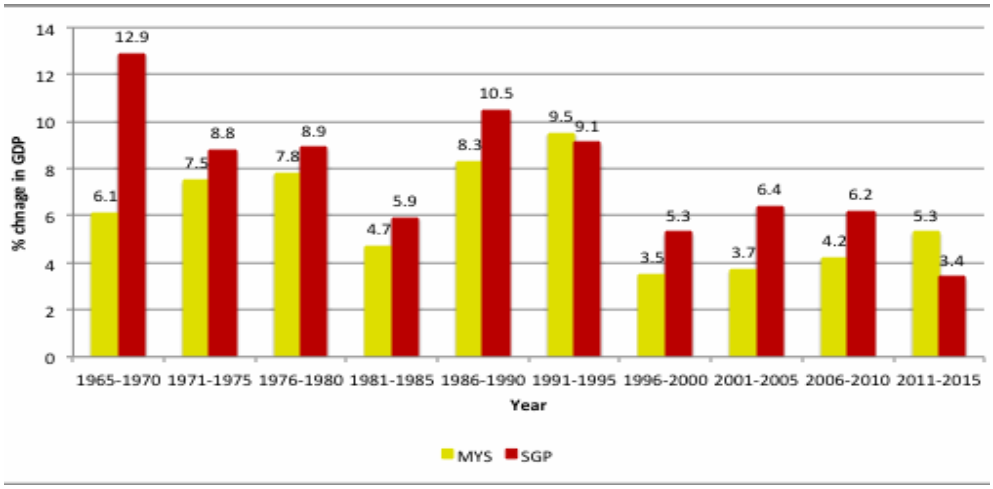
With little money and resources to build on, Singapore focused on developing its manufacturing and service sector after gaining independence. Prime Minister Lee Kuan Yew's wanted to to make Singapore an attractive destination for foreign investment, a policy which he hoped would bring in a top class workforce (Hussain, 2015). Additionally, Singapore also placed an emphasis on developing a superior infrastructure, thereby creating strong connections to the outside world for air travel and sea channels. Coinciding with this

policy was the plan to develop an efficient bureaucracy free of corruption, and a clean and green environment. From the very beginning, Singapore had a vision of how it turns the country into an attractive destination for operating business, thereby explaining in part the near 13% economic growth it had just five years after independence. Barro (1997) has pointed out that many of these policies result in higher growth rates, and thereby higher per capita GDP.

As for Malaysia, its plan following separation was to further develop its primary industry, taking advantage of its wealth in resources such as rubber, oil palm, copper, timber, coal, tin and petroleum. Slow but persistent development in this industry resulted in Malaysia's growth through exporting these resources, but also explains its decline in growth as raw material prices fell, especially in the '80s.

Figure 1 shows real GDP growth for Malaysia and Singapore starting in 1965, the year of separation. From 1965 to 1990, real GDP growth in Singapore exceeded GDP growth in Singapore for every 5-year period, sometimes, as from 1965-1970, by substantial margins. Recall from Table 1 that per capita GDP in 1965 was already 67% higher in Singapore. The subsequent higher growth rates of real GDP, in the absence of substantially different population growth rates, simply widened the gap. The initial growth rate of 12.9% from 1965 to 1970 in Singapore is extraordinary. China, these days, is probably growing at half that pace. If an economy experiences growth of 13% a year, then its output doubles roughly every 5.5 years! Even growth rates of 9%, roughly the average Singapore saw from 1971 to 1995, double real GDP every 8 years, meaning this happened approximately three times over that period!

Figure 1: Real GDP Growth Rates, Malaysia and Singapore, 1965-2015, 5-Year Intervals



Source: *World Bank, GDP in LCU, MYS and SGP*

To illustrate the difference in growth rates further, assume that both countries started at the same point in 1965, say at 100. By 1995, real GDP in Singapore would be 960, while for Malaysia it would be "only" at 576, meaning that Singapore's GDP would be higher by a substantial margin (by 67%). Note, however, a main point of this analysis: Malaysia was not experiencing low growth - indeed real GDP was over five times as high by 1995 when compared to its level at the independence event; instead, it is the extremely

high growth that Singapore experienced that turned its performance into one of the all time winners in that category.

Following the Asian Financial Crisis, which clearly resulted in lower growth rates during the 1996 to 2000 period, but also subsequently, the two countries have performed more similarly, with Malaysia actually increasing real GDP growth rates for every 5-year period to the point where it outpaced the growth rate of Singapore from 2011 to 2015.

Looking at annual growth rates,[45] both countries show a decline for real GDP for the following years: 1985, 1997, 2001, and 2009. Some of these were caused by external events, such as the Asian Financial Crisis in 1997, the World Financial Crisis in 2008-2009, and the dot-com recession coupled with the SARS virus scare. Other declines, such as the one in 1987, were the result of internal policies.

In summary, Singapore started at a higher per capita income level when it became independent and was able to widen the gap, but not as a result of poor performance by Malaysia, which experienced growth rates that many developing countries would be proud of. Instead, the difference in living standards widened as a result of extraordinarily high growth rates in Singapore.

Table 2 further solidifies the conclusion by adding population growth into the equation, and by evaluating GDP growth in USD. Note that, to an approximation, per capita GDP growth equals real GDP growth minus the population growth rate.[46] The data show that Singapore outperformed Malaysia for every period with the exception of 2000 to 2009. The observed differences between Figure 1 and Table 2 must be driven by exchange rate movements, which I will analyze further below.

Table 2: Annual Average GDP, Population, and per capita GDP, Malaysia and Singapore, Selective Periods, US $

	MYS			SGP		
	GDP growth	Population growth	GDP per capita growth	GDP growth	Population growth	GDP per capita growth
1965-1979	15.1	2.5	12.3	17.5	1.7	15.5
1980-1989	**5.3**	2.8	**2.4**	**11.0**	2.2	**8.6**
1990-1999	**6.7**	2.6	**4.1**	**10.1**	3.0	**7.0**
2000-2009	8.9	1.9	6.9	8.1	2.4	5.5
2010-2015	3.0	1.5	1.5	4.4	1.7	2.6

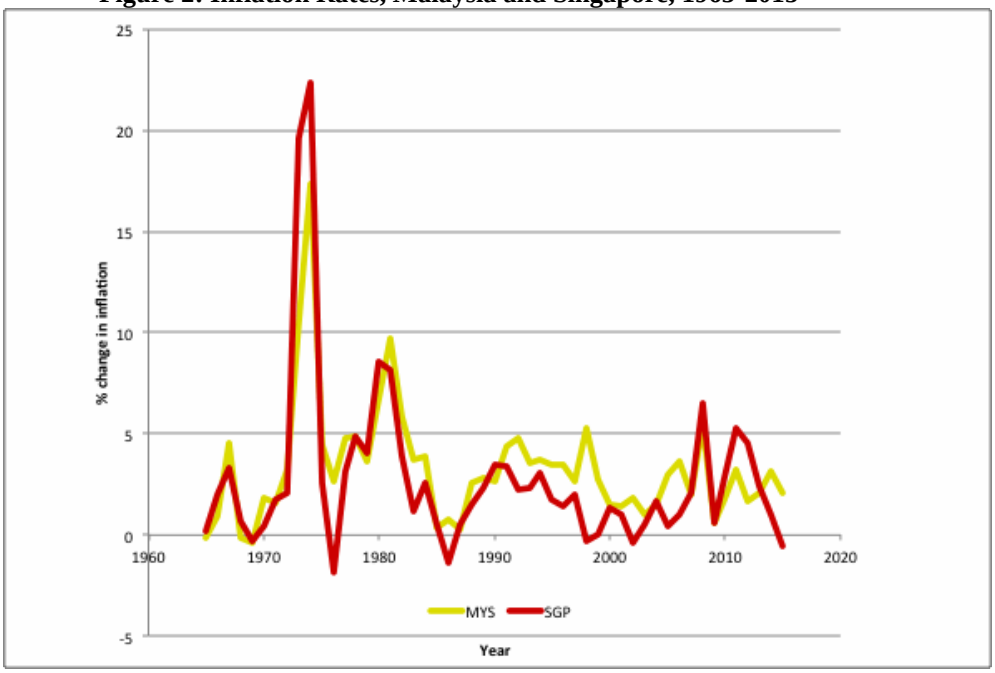
Figure 2 looks at the respective inflation rates for the two countries. The sample period can be divided into two parts. Inflation in both countries was highly volatile from

[45]To keep the length of the paper to a reasonable level, I decided not to include all graphs. These are available from the author by request.

[46]The approximation is based on the growth rate of a ratio equaling the growth rate of the numerator minus the growth rate of the denominator. The resulting numbers are closer for smaller growth rates.

1965 to the early '80s. After that, inflation has typically been below 5% with a few exceptions like Malaysia in 1998 at 5.3% and Singapore in 2009 and 2011 with 6.5% and 5.3%. It averaged close to 2% for both countries since then, although Malaysia saw consistently higher inflation rates than Singapore during the '90s. During the earlier period, the extraordinarily high inflation rates of 22% (Singapore) and 17% (Malaysia) in 1974 stand out. These were the results of the first oil crisis (OPEC I). There was another spike in the early '80s following OPEC II, with inflation rates close to 10% for both countries. Note that there are also several periods of mild deflation, more often seen in Singapore than in Malaysia.

Figure 2: Inflation Rates, Malaysia and Singapore, 1965-2015



Since inflation rates are similar between the two countries, they cannot give us an explanation for the divergence in economic growth from this source. Barro (1997) does not find inflation rates to affect growth rates negatively unless they exceed certain thresholds, which are higher than those experienced in the two countries.

Figure 3 displays the unemployment rate behavior in both countries from 1990 until now. Data are not available for the pre-1990 period unfortunately. In general, unemployment rates are low when compared to Western industrialized countries. Full employment in the U.S., for example, is considered to be around 5% and at the height of the recession reached 10.1%. It is still 23.5% in Greece today, and 19.6% in Spain.

Both countries experienced very low unemployment rates of close to 2% before the onset of the Asian Financial Crisis. The negative economic shock resulting from the Crisis caused Singapore's unemployment rate to be higher than Malaysia's, but both countries have displayed very similar behavior since then.

Figure 3: Unemployment Rate, Singapore and Malaysia, 1991-2014

The similarity of the unemployment rate behavior masks significant differences in labor market policies between the two countries. To gain a better understanding of employment behavior, it is worthwhile to look at some deeper issues in their respective economic policies.

Starting with Malaysia, in 1971, the government announced the *New Economic Policy* (NEP) which essentially aimed to eradicate poverty and restructure the Malaysian society to overcome economic imbalances and inequalities. The policy also tried to restructure the labor market and the distribution of occupations among participants. The hope was to eliminate the so-called 'ethnic structure of labour' which the ruling party believed to have been created by the 'British Imperialist and aggressive Chinese immigrants,' both of which were seen as having marginalized the Malays in their own country. Malaysia's Prime Minister, Dr. Mahathir, argued that the Malays were at a disadvantage in the economic sector and believed they should have higher participation rates since Malays were the majority (Schuman, 2009). Thus, the new policy required Malays (Bumiputera) to have ownership and control 30% of the entire corporate sector. One outcome of the policy was to reduce poverty in Malaysia. In 1970, almost half of all households in Peninsular Malaysia lived in poverty. By 1990, this number was reduced to 17.1% (Snodgrass, 1995, 11). The NEP undoubtedly increased employment in Malaysia as the majority of the population were Malay. Bumiputeras are now represented in a majority of the jobs in the service sector though they are significantly underrepresented in higher level jobs in the private service sector. This outcome is clearly the result of the government's apparent favouritism, offering Malays jobs based on their race rather than being the result of their qualifications. Unfortunately, this exacerbated the segregation of races in Malaysia. Instead of strengthening the society by becoming more equal, it seems to have had the opposite effect.

Singapore took a very different approach in its industrial policy. In 1961 Singapore established its *Economic Development Board* with the primary aim of making the country

an attractive destination for foreign capital. Singapore began the 1960s with high unemployment as its manufacturing sector was underdeveloped, partly the result of its over-dependence on entrepôt trade (Cahyadi *et al*., 2004, 5). To attract foreign investment, the country implemented an incentive scheme consisting of tax benefits which initially were supposed to last up to five years (the time limit was extended subsequently). The intention was to lower production costs for foreign corporations, thereby giving them an incentive to move production to Singapore. The policy had the intended results as more foreign corporations opened operations in Singapore. The government, meanwhile, ensured that the country upheld its high standards of working environment so that it could attract foreign investors to work in Singapore. The 1980s saw upward pressures on workers' wages because Singapore did not want to be perceived as a low-wage country. Consequently, it implemented a high wage policy to incentivie employers to use labour more efficiently and moved into the IT sector to remain competitive. With hindsight, this was a smart move by the government of Singapore as the local market now could meet the wage demands of skilled workers in the high tech industry, curbing their issue of high unemployment of roughly 10% in the 1980s. However, the wage increases soon outpaced the productivity, leading to higher unit labor costs and leading to uncompetitive export prices which resulted in the -1.4% real GDP growth rate in 1985. The fluctuations in unemployment post 1990s is closely linked to ups and downs of the global IT industry: note that the country saw a peak in its unemployment rate in the early 2000s just after the dot-com bubble burst.

Note that similar to many Asian countries, there is not much variation in the unemployment rate, especially when compared to western developed countries. Hence I do not expect the (Expectations Augmented) Phillips Curve to provide a good fit for the data.[47]

Figures 4 (both currencies against the USD) and 5 (Ringgit to Singapore Dollar) show that foreign exchange rates in both countries underwent very different journeys throughout the last fifty years. Malaysia and Singapore stayed at similar exchange rates, $0.33USD to $0.46USD per ringgit (MYR) and Singapore dollar (SGD), from 1965 to 1980. Hence, the ringgit to SGP exchange rate was constant.[48] Unfortunately, Malaysia's exchange rate was in continuous decline until 1997 when the rate plummeted to $0.25USD/MYR. Following the crisis, Malaysia was one of the few countries that decided to be on a fixed exchange rate to the USD for the next eight years. This partly explains why PPP gives such a poor fit for Malaysia; it is only intended to work for countries with a floating exchange rate. According to PPP, a country's exchange rate will depreciate if it has a higher inflation rate than the foreign country. Indeed, since the mid '80s, Malaysia has experienced higher inflation rates than Singapore, as pointed out above while discussing the respective inflation rate behavior (see Figure 2 above). However, I also found that when plotting the percent appreciation/depreciation against the inflation rate differential, the trend line fit was not particularly good, nor did it have the expected slope of one for either country.[49]

[47]Again, I attempted to estimate a Phillips Curve for both countries, fitting a trend line through the change in the inflation rate and the unemployment rate. The results were discouraging, and are available from the author by request.

[48]Exchange rates were fixed until 1972.

[49]Again, to save space, this analysis had to be omitted from the paper but is available from the author by request.

Figure 4: Foreign Exchange Rate, USD vs MYR and SGD, 1965-2015

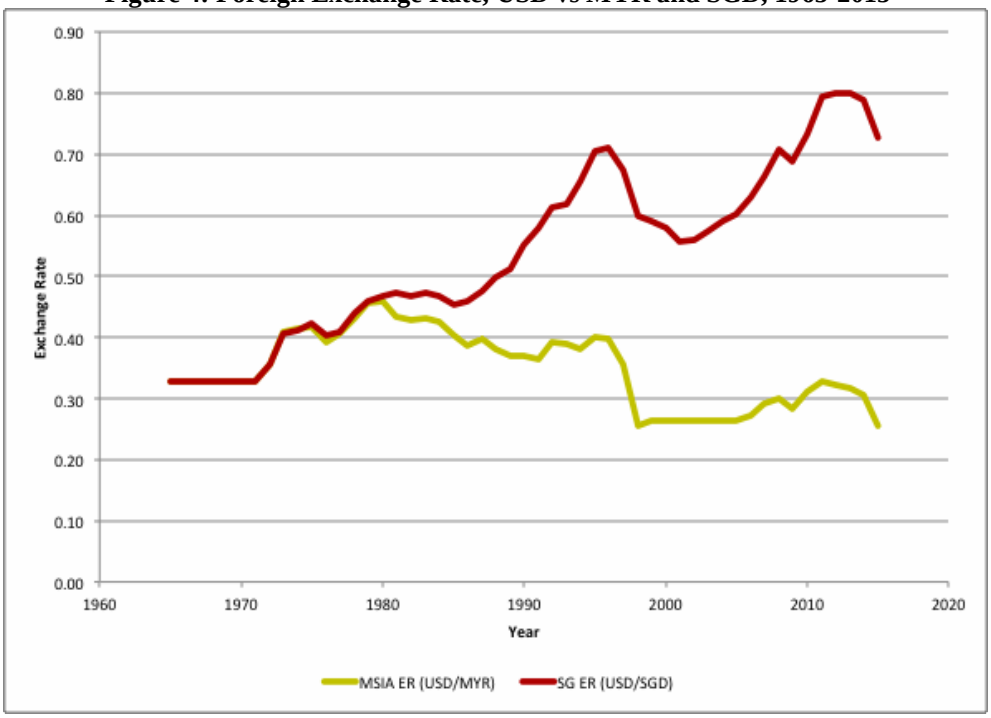
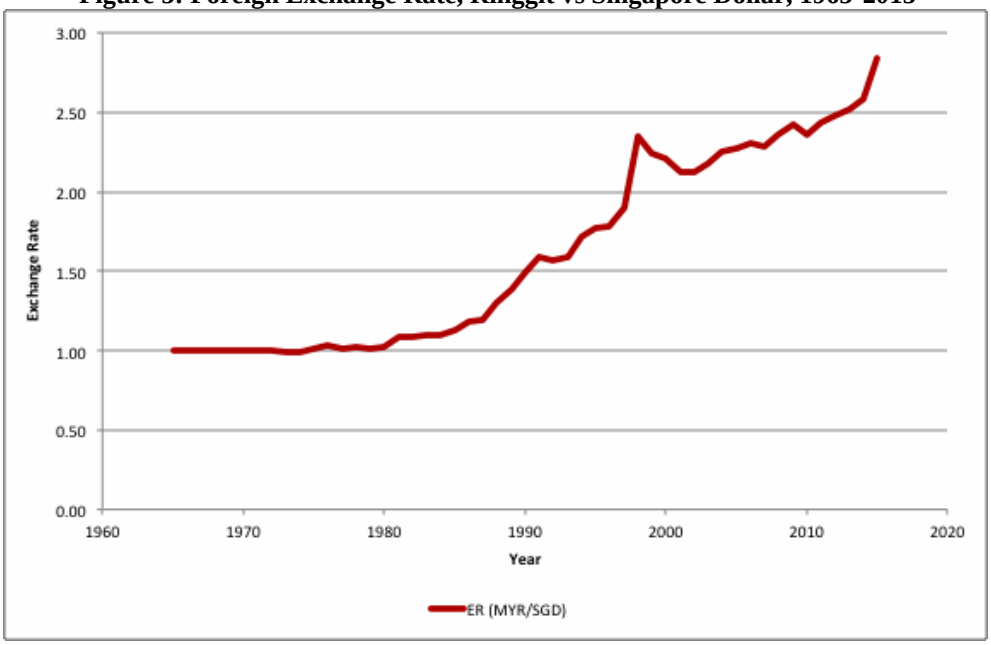


Figure 5: Foreign Exchange Rate, Ringgit vs Singapore Dollar, 1965-2015



Since 2010, the rate has been set on yet another decline, with the rate being at $0.26USD/MYR in 2015. In contrast, the Singapore Dollar has appreciated significantly over time, with its exchange rate reaching to $0.71USD/SGD in 1997.[50] Similar to Malaysia, the Singapore rate experienced a significant depreciation in 1996-7 during the Asian Financial Crisis. Over the course of ten short years, the rate skyrocketed from a trough of $0.56USD/SGD in 2001 to a peak of $0.80USD/SGD by 2011.

One of the reasons for Singapore's long term appreciation is its economic development throughout the years.It evolved from a predominantly low-skill labour intensive workforce to a highly skilled and large capital base workforce to attract foreign businesses. This structural change in the economy has resulted in an influx of FDI which has strengthened its balance of payments to a surplus, giving rise to its strong currency. Over the years, Singapore chose a managed float exchange rate policy, believing it was the best way to achieve its main objective - maintaining the purchasing power of the SGD (Monetary Authority of Singapore, 2001). As well as being an ideal way to deal with external economic shocks and uncertainty such as the Asian Financial Crisis, the managed float also enabled Singapore to sustain its long track record of low inflation rates.

In the early days, Malaysia's economy relied heavily on agriculture export commodities which can explain the volatility of its currency that was heavily dependent on global commodity prices (OANDA, 2016). Initially Malaysia adopted a post 1972 floating exchange rate regime soon after its independence, and switched its 'official currency' from sterling pounds to the US dollar on the foreign exchange market. During the oil crisis of 1973, the U.S. dollar grew unstable causing not only an unstable ringgit but also a spike in Malaysia's inflation rate. Consequently, the central bank decided to switch the country's economic policy to a managed float which meant that governments would intervene in the foreign exchange rate market in an attempt to determine the exchange rate (Talib, 2004). This regime managed to keep Malaysia's exchange rate relatively stable until the Asian Financial Crisis which sent the ringgit in a downward spiral. To prevent further depreciation of the currency, the ringgit was pegged at USD$1.00 = RM3.8010 for six years. The ringgit has experienced tumultuous years since 2010 due to falling oil prices and the 1MDB affair, a major corruption scandal in the country.

After analysing the main economic indicators of well being for both economies, we still do not have a good answer to the question posed at the very start of the paper, namely how did Singapore move from being a lesser developed country to one of the wealthiest nations in such a short time? As indicated by Mankiw (2013) there is an economic theory based on the work of Nobel Prize winner Robert Solow, that tries to determine the per capita income level of countries. In these so-called "growth regressions," a trend line is fitted to a country's growth over longer periods of time on its determinants.[51] I have used the trend line (regression) results from Barro (1997) to predict how differences in underlying variables can explain the levels of development in the two countries.

[50]It is more common to express the exchange rate in ringgit/USD and Singapore Dollar/USD. The latest numbers are 4.14 for Malaysia and 1.40 for Singapore. However, the advantage of listing the inverse in my paper lies in an increase in the ratio indicating an appreciation for the home currency (Malaysia and Singapore relative to the USD), while a decrease stands for a depreciation. For the ringgit to Singapore dollar I followed the standard and listed the ratio that is greater than one, implying an appreciation for the Singapore dollar if the ratio increases.

[51]A subtle point is that the explanatory variables include the initial level of per capita GDP. Coupled with a growth rate over a subsequent period. This results in the determinant of the current per capita GDP level. I used Keil (2015) lecture notes in Intermediate Macroeconomics for an initial understanding.

Barro's results indicate that, controlling for other effects, there is a "catch-up" effect, meaning that, ceteris paribus, countries that are further behind will grow faster (*initial level effect*). This would favor Malaysia over Singapore, since Singapore was more developed in 1965 as its per capita GDP level indicates.

Population growth plays a crucial role in the Solow Growth Model, with higher population rates resulting in lower growth: The fertility rate in Malaysia has always been larger than in Singapore. For example, the Malaysia to Singapore fertility rate was 4.9/3.4 during the 1960-1980 period, and even though fertility rates in both countries have fallen, it was still twice as high in Malaysia for the years 2001-2014 (2.2/1.3).

The *democracy index* plays a two-edged role in Barro's model: there is a benefit initially as a country comes out of dictatorship, but this flattens out and even becomes a negative influence as special interest groups seem to influence a country's economic policy. Note that Hong Kong was never a democracy but that India has a long tradition in it. Regardless, the Democracy index is not available for Malaysia and Singapore for the 1960-2000 period, and when it was published for the 2000-2014 sample period, it was very similar for the two countries.

The *Government Consumption ratio*, or the share of government expenditures in GDP, has a negative effect on growth in Barro's work. Here Singapore is the winner, since its government consumption ratio is lower throughout the sample period, with Malaysia's government expenditure ratio of 12-16% compared to Singapore's 9-10%.

Barro's work stresses *education*, in particular male secondary and higher schooling experience.[52] Education enters his equation in two ways: first of all, higher education results in faster growth in itself; but in addition, it also leads to a faster convergence. Data for male secondary and higher schooling is somewhat inaccurate as only data from several years were available. On the other hand, I do not expect large year-to-year fluctuation in that variable. According to Barro, the higher the value of males receiving more advanced education, the higher the economic growth. The data shows that Malaysia and Singapore both have similar percentages of students progressing to secondary education although Singapore has an edge here with a higher level of male education (96.8 versus 73.8 for the 2000-2014 time period).

Perhaps the starkest difference between the two countries lies in the "*Rule-of-Law Index*" which Barro uses as an explanatory factor. The index is only available for both countries since 1980, but Singapore's is clearly higher than Malaysia's here (1980-2000: 1.3 versus 0.5; 2000-2014: 1.7 versus 0.5). The final variable considered by Barro also gives Singapore an edge: it is *life expectancy*, which has increased for both countries since 1960 but typically has been four years higher for Singapore, reflecting a work force that can work for a longer period of time, increasing productivity.

Based on this analysis, we should have expected Singapore to grow faster over the 1990-2014 period. The results of our analysis of the Solow Growth Model as applied by Barro (1997) for Malaysia and Singapore gives Singapore an advantage in almost every factor. Singapore has a lower population growth rate, a higher quality of life (life

[52]Again, there is a subtlety here. It is not that female education does not matter, but it already is captured in the fertility rate, which depends on female education. Lower fertility rates indicate that more females are getting a higher education as more educated women tend to have smaller families due to their involvement in the workforce.
To clarify, female education does not contribute to the explanation of growth above and beyond the effect of lowering fertility rates.

expectancy), a lower government consumption ratio, higher education rates, and a higher value for the "Rule-of-Law."

IV. Conclusion

In the past half century, economic development has been at the forefront of Malaysia's and Singapore's growth as both countries quickly broke through to become newly developed and first world countries respectively. Singapore brilliantly invested in technological and human capital in order to overcome their major disadvantage as a small country with little to no resources. Former Singaporean prime minister, Lee Kuan Yew, and his deputy prime minister at the time, Goh Keng Swee, made wise decisions that will forever be at the root of Singapore's future developments. These included plans to transform the country into an attractive business hub that would draw foreign investment which subsequently drove the country's success. On the other hand, Malaysia did well in making use of its natural resources as a way of gaining revenue and forming strong ties with other countries. Prime Minister Mahathir Mohammad's New Economic Policy was successful in drawing the majority of Malays into the workforce. However, the policy also contributed to inequality as it was said to have been biased towards the Malays. Perhaps the reason Malaysia did not thrive as well as Singapore did was because of the corruption in the country, which has only skyrocketed in recent years.

My analysis of the main economic indicators of well being throughout this paper has led me to make the following predictions for the economy of both countries in the next five years. In terms of real GDP growth rate, both countries will continue its trend of approximately 2.5-5.5% per annum. I expect Malaysia's inflation rate to increase due to its falling exchange rate and Singapore's inflation rate to continue to be at a negative level or perhaps slight inflation as its economy undergoes a period of recession this year (2016). Singapore's unemployment rate is poised to rise to over 3% meanwhile, Malaysia should anticipate a decline in its unemployment rate. However, this is heavily dependent on the recovery of the world's economies- such as the United States, United Kingdom and China- that provide support to Malaysia's export-based economy. As for the exchange rates, I foresee a decline in Malaysia's exchange rate to an all time low of $0.21USD/MYR and the strengthening of Singapore's exchange rate to $0.66USD/SGD in the near future.

Future research should explore a more detailed analysis of the differences in institutions that generate the differences in the driving variables for the Solow Growth Model. This would enable us to answer questions like why is female fertility still higher in Malaysia?; why is law and order lower in Malaysia?; and what causes the lower level of education in Malaysia? Another possibility for research would be comparing Malaysia with one of the lesser performing Southeast Asian countries like Thailand or Indonesia.

V. Bibliography

Abisheganadan, F. (1965). "Singapore Is out." News Straits Times, 1965.

Barro, R. (1997). *Determinants of Growth: A Cross-country Empirical Study. Cambridge*, Massachusetts: MIT Press.

Bank Negara Malaysia (2015). "BNM Financial Market." May 27. Accessed September 1, 2016. http://www.bnm.gov.my/index.php?ch=en_fxmm_mo&pg=en_fxmm_mo_overview&ac=453&lang=en.

Cahyadi, G., Kursten, B., Weiss, M, and G. Yang (2004). "Singapore's Economic Transformation." Global Urban Development: Singapore Metropolitan Economic Strategy Report, June. Accessed September 20, 2016. http://www.globalurbandevelopment.org/GUD%20Singapore%20MES%20Report.pdf.

Cheah, B. (2002): "The Making of a Nation." Singapore: Institute of Southeast Asian Studies.

OANDA Corporation (1996). "Malaysian Ringgit." Accessed September 2, 2016. https://www.oanda.com/currency/iso-currency-codes/MYR.

Department Of Statistics Malaysia (2015) "Current Population Estimates, Malaysia, 2014 - 2016." Accessed August 20, 2016. https://www.statistics.gov.my/index.php?r=column/cthemeByCat&cat=155&bul_id=OWlxdEVoYlJCS0hUZzJyRUcvZEYxZz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09.

Department Of Statistics Singapore. (2016) "Statistics Singapore." Accessed August 20, 2016. http://www.singstat.gov.sg/statistics/latest-data.

Department Of Statistics Singapore. (2015) "Population Trends- 2015." Accessed August 25, 2016. https://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/population_and_population_structure/population2015.pdf.

Fong, L. (1990) "The Week Before Separation." News Straits Times

Huff, W. G. (1999) "Singapore's Economic Development: Four Lessons and Some Doubts." Oxford Development Studies 27, no. 1 (February 1999): 33–55. doi:10.1080/13600819908424165.

Hussain, Z. (2015) "How Lee Kuan Yew Engineered Singapore's Economic Miracle." BBC Business (BBC News) Accessed September 20 2016. http://www.bbc.com/news/business-32028693.

IMF World Economic Outlook Database List (2016) "Information about Gross Domestic Product (GDP)." Accessed September 20, 2016. http://www.imf.org/external/ns/cs.aspx?id=28.

LLC, Macrotrends. (2010) "Crude Oil Prices - Live, Daily & Historical Charts." Accessed August 30, 2016. http://www.macrotrends.net/1369/crude-oil-price-history-chart.

Index Mundi. (2014) "Malaysia Demographics Profile 2014" Accessed August 25, 2016. http://www.indexmundi.com/malaysia/demographics_profile.html.

Keil, M. (2015) "The Solow Growth Model" *Lecture Notes*, Intermediate Macroeconomics, Claremont McKenna College

The Commonwealth (2016) "Malaysia: History." Accessed August 25, 2016. http://thecommonwealth.org/our-member-countries/malaysia/history.

Monetary Authority of Singapore. (2001) "SINGAPORE'S EXCHANGE RATE POLICY." Accessed September 2, 2016. http://www.mas.gov.sg/~/media/manual%20migration/Monographs/exchangePolicy.pdf.

Nathan, K. S. (2002) "Malaysia–Singapore Relations: Retrospect and Prospect." Contemporary Southeast Asia 24, no. 2 (August 2002): 385–410. doi:10.1355/cs24-2i.

Country Studies. (2003) "Singapore - Road to Independence." Accessed August 20, 2016. http://countrystudies.us/singapore/10.htm.

Singh, D. and Arasu V T. (1984) "Singapore, an Illustrated History, 1941-1984." 2nd ed. Singapore: Information Division, Ministry of Culture.

Snodgrass, D. "Successful Economic Development in a Multi-Ethnic Society: The Malaysian Case." Accessed September 20, 2016. http://earth.columbia.edu/sitefiles/file/about/director/pubs/503.pdf.

Talib A.L. (2004). "Pegging The Ringgit Against The US Dollar: An Evaluation" Accessed September 20, 2016. https://www.statistics.gov.my/portaL_@Old/download_journals/files/2004/Volume1/Contents_Article_abdul.pdf

Tang, C.F. (2009) "The Linkages Among Inflation, Unemployment and Crime Rates in Malaysia." Int. Journal of Economics and Management 3, no. 1 (2009): 50–61. Accessed September 10, 2016. http://econ.upm.edu.my/ijem/vol3no1/bab04.pdf.

The Economist. (2013) "The Economist Explains." Accessed September 19, 2016. http://www.economist.com/blogs/economist-explains/2015/03/economist-explains-23.

Turnbull, C M. (2009) "A History of Modern Singapore, 1819-2005." Singapore, Singapore: Singapore : NUS Press, c2009.

Umezaki, S. (2006) "Monetary and Exchange Rate Policy in Malaysia Before the Asian Crisis." INSTITUTE OF DEVELOPING ECONOMIES DISCUSSION PAPER No. 79 (2006). Accessed September 15, 2016. http://www.ide.go.jp/English/Publish/Download/Dp/pdf/079.pdf

Lee, K.Y. (1998) "The Singapore Story: Memoirs of Lee Kuan Yew. 4th ed. London, United Kingdom: Pearson Education Imports: Depositories, 1998.

World Bank (2016) "GDP Growth "http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG Accessed August 20 2016.

World Bank (2016) "Total Population" http://data.worldbank.org/indicator/SP.POP.TOTL Accessed August 20 2016.

The Silver Lining Behind the Darkness: Social Media as an Innovative Tool to Combat Sex Trafficking in Southeast Asia

Yutong Huang

Author background: Yutong Huang grew up in China and currently attends The Affiliated High School of South China Normal University, located in Guangzhou, China. Her Pioneer seminar topic was in the field of culture studies and titled "Globalization and International Migration."

I. Introduction

Kevin Bales, a prestigious scholar who studies contemporary slavery, once said: "If there is a fundamental violation of our human dignity that we would all say is horrific, it's slavery" (Bales).

What happens to people who are trafficked? They become slaves.

United Nations' Universal Declaration of Human Rights (UDHR) strongly advocates for the freedom, a fundamental right of all people. In Article 3, 4 and 5, UDHR upholds people's "right to life, liberty and property," denounces all forms of slavery, servitude, and cruel treatments to human ("The Universal Declaration," pars. 11 - 13). Yet according to the International Labor Organization, approximately 20.9 million people are victims of human trafficking globally ("The Victims," par. 1). The precise statistics and figures about human trafficking remain unclear, though, due to the fact that the trafficking is illegal, and, therefore, "invisible" ("Human Trafficking," par. 4). However, human trafficking is inarguably one of the major violations of human rights, and has, over past decades, increasingly gained the attention of the international community. Among the subgroups of human trafficking, sex trafficking no doubt receives much of the attention because of its brutal exploitation of vulnerable young women and children. The victims of sex trafficking have diverse backgrounds in terms of social class, economic ability, education and age. While the victims span all sorts of demographics, there are indeed some vulnerabilities that lead to a higher susceptibility to victimization of sex trafficking: runaways and homeless youth who normally have not received high education ("The Victims," par. 5), foreign nationals who do not enjoy the same rights as citizens do (par. 6), and even ordinary poor country girls living in the rural areas are likely to be trafficked, because they are inexperienced and believe the deceitful promises of traffickers who take advantage of the victims' vulnerabilities, and force them into commercial sex.

Southeast Asia is the region where sex trafficking is most prevalent. Children in Cambodia, Thailand, Vietnam and other Southeast Asian countries, often as young as four and from poor families that are desperate for money, are sold to traffickers to supplement the family income. Their families are told that their children will be employed in respectable

places and will be able to mail remittances back home. But what their families do not know is that their children are actually sold to brothels, beaten, ill-treated, drugged, raped, and eventually, prostituted ("Child Sex," par. 1). Prostitution in Southeast Asia even developed into a thriving "tourist industry" — sex tourism — in which foreign tourists come and "enjoy" the young flesh. Such violent abuse not only brings substantial damage to the victims' physical health, but also causes permanent psychological trauma. Therefore, solutions are needed to combat sex trafficking and save millions of young boys and girls from suffering.

The international community, national and local governments as well as non-government organizations, are devoting significant resources to combating sex trafficking. All countries in Southeast Asia have signed the UN declaration of Human Rights in 1948 to promote the rights of "life, liberty and property" of their citizens. In addition, ten member states of the Association of Southeast Asian Nations (ASEAN) have already formed a close cooperation to tackle the issue of human trafficking, which has resulted in high-level initiatives ("Southeast Asia's," par. 12). To date, all member countries of ASEAN, except Laos and Singapore, have passed anti-trafficking legislation (par. 20), and all ASEAN countries are part of the Bali Process, the International Organization for Migration, against trafficking in persons, smuggling, and related crimes (par. 18). Each member country also has different levels of measures taken against trafficking in persons. Yet in practice, the current combat effort is barely working, due to the complexity of the trafficking industry, such as the fact that the police are often protecting the traffickers. As the agency that is supposed to protect the people facilitates the exploitation, human and sex trafficking still remain a serious problem for every level of law enforcement. As a result, it is time for us to invent new wheels to deal with the issue — creative thinking needs to be called upon. With the prevalence of Internet access and cell phones, social media are potentially powerful tools to combat sex trafficking, given its popularity and accessibility.

Social media refers to social networking platforms like Facebook and Twitter, advertisements, and apps. In this paper, I will explore the extent to which and how social media could be used to help in the struggle against sex trafficking from a human rights perspective. In other words, this paper will answer the question of how social media can be used to combat trafficking and protect the human rights of vulnerable women and children. By looking at each of a few typical case studies demonstrating how social media are used in preventing and detecting sex trafficking and in protecting the survivors, I will propose actionable measures for the use of social media while also acknowledging that social media is indeed not the only tool to solve sex trafficking but one of the tools that can help. Along the paper, I will not only give practical recommendations for social media in accordance with the 3P paradigm, but also address limitations that should be taken into consideration when implementing.

From a human rights perspective, this paper aims to provide a valuable insight into how society can use social media, a comparatively unexplored field in traditional sex trafficking combat, as a weapon to fight sex trafficking and to protect millions from heading onto the wrong path. The application of measures in social media can also be extended to combatting inequality in gender, for the socially marginalized, and in human rights in general.

By the end of this paper, readers should have a clear understanding of my research findings, including the following: basic information about sex trafficking in the global and regional (Southeast Asia) context, current campaigns against sex trafficking from governments, inter-government organizations, non-government organizations, civil societies on the international and national levels, and most importantly, a practical yet novel view into how to use social media to fight sex trafficking in Southeast Asia.

II. An Overview of Sex Trafficking

A. A Global Perspective of Sex Trafficking

Human trafficking is a form of modern slavery. It is also one of the world's most serious and shameful crimes (see figure 1). By 2012, human trafficking had been reported to deny the freedom of approximately 20.9 million people worldwide ("ILO Global," 13). Due to the complexity of measurement, there is still a large hidden population of victims, and the actual statistics may be even higher. In fact, according to Bradley Myles, deputy director of the Polaris Project, one of the world's most active anti-trafficking organizations, human trafficking is the third-largest and the fastest-growing criminal industry in the world (qtd. in Couch, par. 8).

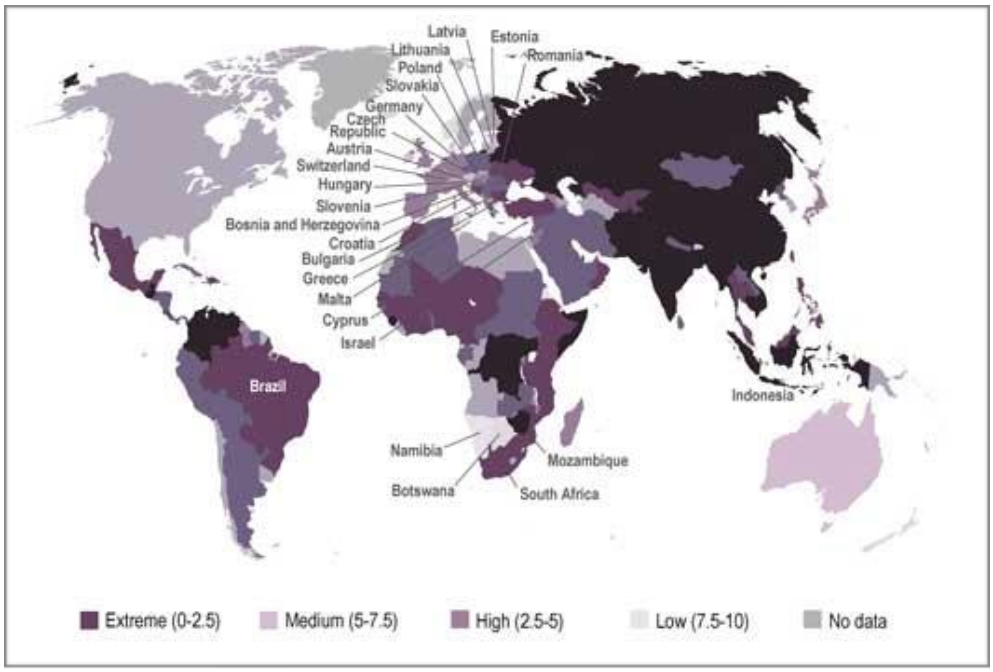


Figure 1: *A Global Map of Human Trafficking Sites (Source: Maplecroft.com)*

One reason for the huge number of victims is that human trafficking comes in many forms. It is officially defined as "the recruitment, transport, transfer, harboring or receipt of a person by such means as threat or use of force or other forms of coercion, of abduction, of fraud or deception for the purpose of exploitation" by *Protocol to Prevent, Suppress and Punish Trafficking in Persons* (qtd in. "Human Trafficking," par. 2)**.** Breaking

down the definition given in the *Trafficking in Persons Protocol* to give a more well-rounded definition, human trafficking must have the following three constituents: the act of recruiting, transferring, and harboring of persons; the use of threats, fraud, violence, coercion, and any other means to keep the persons under control, and the purpose of trafficking (par. 3). Human trafficking significantly violates human rights and was considered by Kofi A. Annan, the former Secretary-General of United Nations, one of the most egregious violations of human rights that the world had ever confronted (Annan iv).

Human trafficking that involves commercial sex acts is called sex trafficking. The International Labour Organization estimated in 2012 that among the identified victims of human trafficking, 22% of them, which is about 4.5 million people, are trafficked for sexual exploitation globally ("ILO Global," 14). As a subset of human trafficking, sex trafficking no doubt receives much attention because of its brutal sexually abusive and exploitative nature and the exceptional vulnerability of its victims, children and young women. Sex traffickers, like any other human traffickers, compel victims into the sex industry against their will — the forms of coercion include but are not limited to fraud, threats, and violence ("Sex Trafficking," pars. 1-2). But the traffickers do not traffic these boys, girls, women and men just to be mean to them. They do it for profit.

Not surprisingly, the ILO also estimated that two-thirds of the yearly illegal profit made from human trafficking, which was about $99 billion, was attributed to sex trafficking ("ILO Profits" 15), making it one of the most profitable criminal acts next to the drug trade and the illegal arms trade. Once delivered to the brothel, the victims' pictures will be listed on the menu for men to choose from, which further objectifies them as commodities and strips them of their humanity. Then they are forced to engage in sexual intercourse — a girl's virginity can be sold for as high as $200 (higher in tourist areas)—while "ordinary" sexual transactions may be sold for as little as $5. The children might be raped between five to ten times a night (Harden, pars. 1-7). If people think that the victims can make a considerable amount of money out of it, they are completely wrong. These victims become slaves as soon as they are trafficked. The truth is that most of the profit flows into the pimps' hands.

Normally, not only do the victims of sex trafficking have relatively low economic status, they are also socially marginalized. These people are particularly vulnerable to trafficking since they are poor, often illiterate, from low-status castes, rural communities or ethnic minorities, intellectually or physically disabled or children born of rape or out of wedlock. In many instances, girls from poor families are looking for jobs to augment their family's household income; traffickers approach them and promise them working opportunities far away from home, which, in fact, are life-changing traps. If the victims ever manage to run away, the traffickers, who withhold their important documents like visas or passports, would immediately turn them over to the authorities and the victims would be arrested as undocumented migrants (Kabance 18). Yet due to the high demand for sex workers, and the relatively high profits compared to the low risk, more people are engaging in this criminal act despite constant struggle against it. While this is an age of globalization, which makes the crime even more sophisticated than ever, there is still a lot of hidden information yet to be collected. Sex trafficking remains a serious issue globally and a difficult offense to deal with. Therefore, combat requires effort on all sorts of levels.

B. Sex Trafficking in Southeast Asia

Turning to Southeast Asia now, sadly, we see a hub of sex-trafficking. Southeast Asia is the most common origin, transit and destination of sex trafficking victims, where countries involved deeply in sex trafficking are Vietnam, Malaysia, Laos, Thailand and Myanmar. Specifically, Thailand, Laos, and Cambodia are countries of origin, transit, and destination, while Vietnam is a country of origin and destination ("2016 Trafficking" 113, 119, 238, 364).

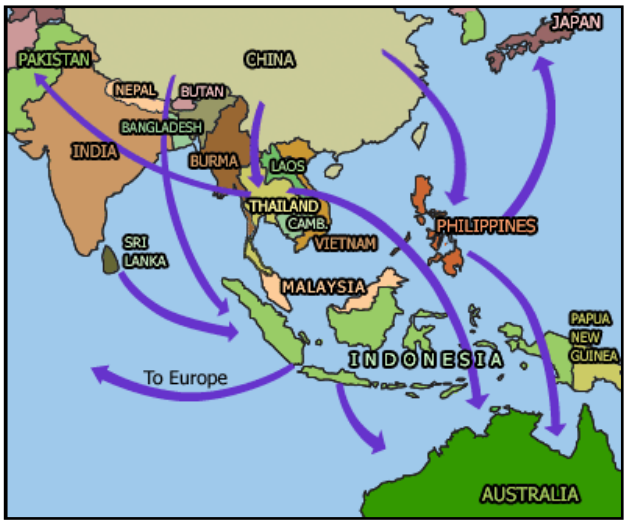


Figure 2: *A Map of Human Trafficking Routes Featuring Southeast Asia and its Neighbors (Source: Public Broadcasting service.org)*

Accounting for 26% of the total trafficked population in East Asia, South Asia and Pacific (see figure 3) ("Global Report" 34), sexually abused victims do not just come from the Asia-Pacific region but are brought to, and are possibly being transferred to, other more affluent places like North America. But for the purpose of this paper, I will mainly focus my recommendations on trans-regional trafficking within Southeast Asia, as it is one of the main forms of exploitation in Southeast Asia (see Figure 4) ("Global Report" 43).

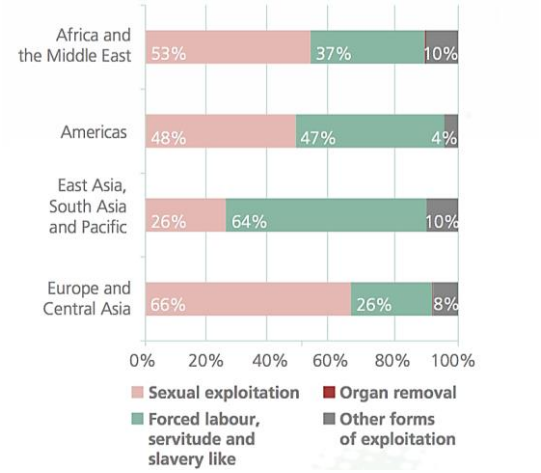


Figure 3: *Forms of exploitation among detected trafficking victims, by region, 2010-2012 (Source: UNODC)*

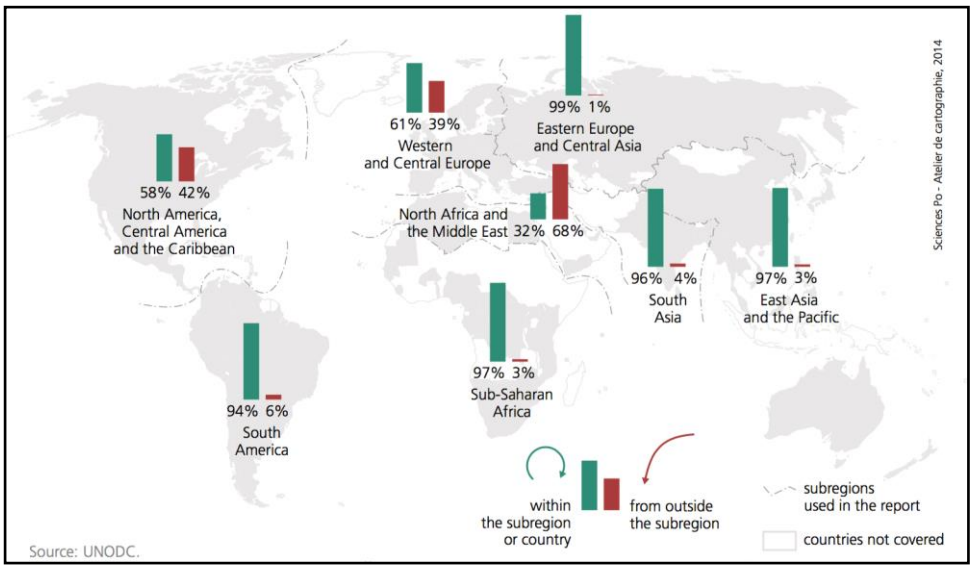


Figure 4: *Shares of detected victims by subregional and trans-regional trafficking, 2010-2012*
(Source: UNODC)

The shocking fact is that Southeast Asia is also a major destination for the crime of child sex tourism (there is, of course, adult sex tourism as well) where rich clients travel for the purpose of sex with minors. For the purposes of this paper, it is important to remember that victims are most frequently trafficked domestically or subregionally. In other words, victims are shipped involuntarily to neighboring countries and within national borders. Looking at the demographics of people who are convicted of trafficking, almost every trafficker has local citizenship (about 97%), committing crimes within their own country, while only 3% of them are foreigners ("ILO Global" 77).

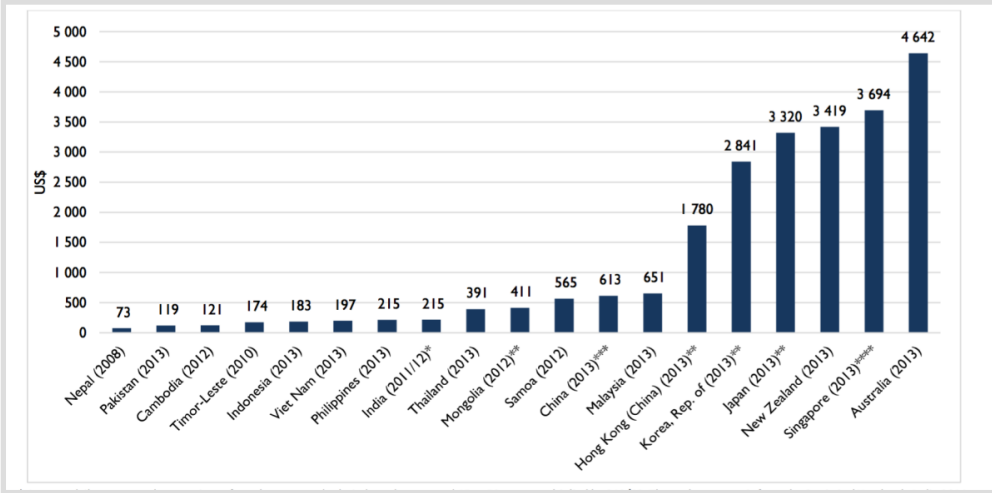


Figure 5: *Average monthly wages in countries from Asia and the Pacific with broadly comparable data, 2013 or latest available year (US$)*
(Source: ILO Regional Office for Asia and the Pacific)

People may wonder why sex trafficking it is so prevalent in Southeast Asia in particular. The prevalence of sex tourism in mainland Southeast Asia can be explained by social and economic factors. Unlike prosperous countries like Brunei and Singapore where women are in a comparatively favorable position because of their relative economic autonomy, Southeast Asian countries like Cambodia, Vietnam and Laos are some of the most impoverished countries in the world. The average monthly wages in Cambodia is only $121 and $197 in Vietnam (see figure 5) ("Wages in" 2).

Socially, women are not in a good position in these traditional, patriarchal societies. Daughters are considered "others' possessions" since when they are married, they normally blend into their husbands' families; sons, on the other hand, are generally given much more worth since they shall carry on their family names. Understandably, then, girls are more easily trapped in sex tourism than boys because they receive less attention and less education. So even if they are lured into sex tourism, often their families will not bother to report it to the police. Economically, the pull factors play a large role along with the social factor: under the circumstance that daughters need to make money for their families, traffickers entice destitute girls with promises of bright prospects, schooling, training, allegedly well-paid jobs in factories, or even marriage. Worse, often their parents are so poor that they directly sell their daughters to the traffickers. The high demand for cheap sex is so great that this factor also plays along to boost sex trafficking (Kabance 12). The lucrative nature of the sex industry also makes it tempting for the traffickers to commit such crimes: according to INTERPOL, the yearly income that a woman can make in the sex trade for her owners is between $75,000 to $250,000. Yet children can often make even more money because their virginity may be sold at a high price. For example, video evidence of children in Cambodia reveals that one girl may sell for $30, two girls for $60.45 (Cotter 497).

C. A Sketch of Current Global and National Campaigns against Sex Trafficking

Now that the world is paying attention to sex trafficking, organizations and governments have already taken initiatives to fight this crime in collaborations, from intergovernmental organizations like UNODC, UN.GIFT, the Palermo Convention, and European Union, to non-government organizations like the Council of Europe, the Polaris Project, the Nomi Network, and Human Rights Watch.

In the global context, the United Nations is spearheading the campaign: The United Nations Global Initiative to Fight Human Trafficking (UN.GIFT) was launched in 2007 to promote the global fight against human trafficking, and is working with all kinds of stakeholders — governments, business, academia, civil society and the media — to ensure and aid each other's work. In addition to UN.GIFT, the United Nations has also issued the Protocol to Prevent, Suppress and Punish Trafficking in Persons especially Women and Children, which 140 parties have signed to date. The Trafficking protocol, along with the Palermo Convention, have put much effort into ensuring protection and support to those who fall victim ("UN.GIFT," pars. 1 - 2). On the other hand, the Polaris Project, a US-based global civil society, has successfully reported and responded to thousands of human trafficking cases all over the world ("Success," par 1). Other non-government organizations like Human Rights Watch, Nomi Network, and Human Trafficking Organization have also devoted themselves to the struggle.

There are also regional efforts. In Europe, the Council of Europe, the continent's leading human rights organization, is working very hard to try to eradicate trafficking within Europe, especially Eastern Europe. The Council of Europe has adopted the Convention on Action against Trafficking in Human Beings since 2005, with a series of activities and campaigns combating human trafficking. The Convention comprehensively encompasses many kinds of trafficking, whether it be national or transnational, labor or sex trafficking. Its "main added value," according to it description, is on victim protection. For example, it protects victims' rights by providing them with renewable residence permits, and it gives victims, no matter what nationality, a recovery and reflection period of at least 30 days ("About the Convention," pars. 1 - 3).

United States has offered a "3P" paradigm — prevention, prosecution, protection — in the 2000 Victims of Trafficking and Violence Prevention Act (TVPA) to the international world to fight against not only sex trafficking but also other kinds of human trafficking in general ("3Ps," par. 1). This "3P" paradigm has a threefold purpose in dealing with human trafficking: to prevent exploitation, to prosecute the offenders, and to protect the victims. Since then, a number of acts and law have been enforced to ensure the combat of human trafficking, such as the Victims of Trafficking and Violence Protection Act of 2000 (VTVPA), the Trafficking Victims Protection Reauthorization Act (TVPRA) in 2003 and PROTECT ACT in the same year (par. 1). Specifically, the US Congress passed VTVPA to provide T visas to victims of trafficking — so that the victims can have legal documents to temporarily remain in the US ("T Visa," par. 2). Such methods shows a humanitarian side of law enforcement, and have made significant progress combatting trafficking. Additionally, the US Department of State also annually releases a TIP Report (Trafficking in Persons Report) that studies efforts against trafficking by persons, governments, and civil societies across the globe. By compiling a global account on anti-

trafficking initiatives, the TIP Report works to enlighten individuals around the world who care about such matters, mobilizing the whole world to eradicate this horrific crime ("2016 Trafficking" 4).

As in Southeast Asia, governments are forming multilateral cooperation initiatives to deal with the problem. Different NGOs, civil societies, and individual activists are also taking actions. The preceding is part is a brief sketch of the current campaigns against sex trafficking. The next section discusses Southeast Asia more in-depth: Current Effort against sex trafficking.

III. Current Efforts against Sex Trafficking

As introduced in section II, governments and organizations around the globe are taking action in the struggle to eliminate sex trafficking. This section will discuss efforts made in Southeast Asia more specifically, along with elaboration on the global effort.

A. International NGOs and Civil Societies Efforts

Non-government organizations play an important role in combating sex trafficking, since they can act more freely than governments. The Global Initiative to Fight Human Trafficking (UN.GIFT), as mentioned in section II, entered the global fight against human trafficking in 2007. Under multi-cooperation with the International Labour Organization (ILO), the International Organization for Migration (IOM), and other organizations, UN.GIFT aims to decrease the demand for human exploitation, reduce potential victims' vulnerability, provide adequate protection and support for victims, and to ensure the prosecution of criminals. By educating people about trafficking in person, UN.GIFT can hopefully raise awareness in people about how detrimental this it is to humans and society ("The Global," pars. 3-4), and in this way contribute to the cause of human rights protection. By now UN.GIFT has launched a series of forums and seminars — Vienna Forum and Seminar in Brasilia — to discuss effective ways to raise public awareness, tackle the complex issues of recruitments and transportation, and establish partnerships in tackling the problem ("UN.GIFT," pars. 3-10). The UN Trafficking Protocol mentioned in section II also serves to form inter-party cooperation and initiate efforts against trafficking in persons. These two efforts by the UN are more oriented toward prevention and prosecution since they are inter-country collaborations and remain on theoretical levels.

On the other hand, the Polaris Project, a non-government organization, seeks practical ways to support victims trapped in trafficking. Their work includes providing shelter, medical care, and counseling services for survivors of trafficking ("Survivor Support," pars. 2-3). The Polaris Project has also initiated a Global Safety Net, by which the project develops cross-border collaborations to shut down traffickers and, thereby, enhances the safety of victims. It is worth mentioning that the project conducts hotline training in Southeast Asia to facilitate the effort — specifically, Polaris is now "hosting the first bilateral consultation between hotlines in Thailand and Vietnam," the website states ("Global Safety Net," pars. 2-4).

In Southeast Asia, regional efforts have also been active. The Development and Education Program for Daughters & Communities (DEPDC) is a non-government

organization in Thailand. Based in Chiang Rai, this organization dedicates itself to the prevention stage in the "3P" paradigm. It adopts a number of different strategies to inform parents about the potential dangers of illegal sex trade, and hopes this will convince them to educate their daughters and to be more cautious about sex trafficking. The information includes information about contagious diseases like HIV, the undesirable conditions in brothels, and the importance of education. Numerous successes indicate that the organization's arguments have persuaded parents to give more weight to the children's education than to their desire for money ("Prevention," par. 7). Similarly, GABRIELA in the Philippines and the Human Rights Commission in Cambodia are NGOs that actively engage in public awareness campaigns, conducting research, organizing workshops, and advocating for the implementation of sex trafficking laws (par. 8-9). When it comes to the protection phase, NGOs like EMPOWER and Protection of Children's Rights (CPCR) will carry out the support for victimized women and children sex workers to help them recover from the traumas, providing counseling services, educational opportunities, and social welfare to those who have been sexually exploited (Cotter). Last but not least, the Nomi Network, a women-empowerment organization based in Asia, dedicates itself to the protection of victims of trafficking. It puts much effort into rebuilding the survivors' life by providing them with sustainable job positions and technical assistance at work, bridging the gap between these vulnerable workforces with the market demand, and successfully empowering survivors to be economically independent ("Program Model," par. 4).

B. Law Enforcements on the National Level

While there has been much work done in law enforcements against sex trafficking globally, Southeast Asia also made its first significant step as a whole. On November 29th, 2004, ASEAN countries signed the *Declaration Against Trafficking in Persons Particularly Women and Children*, recognizing the severity of the problem of human sex trafficking, and expressing their willingness to enter this battlefield ("ASEAN Declaration," pars. 1 - 19).

Thailand is in the vanguard among Southeast Asian countries acting against sexual exploitation. National Policy and Plan of Action for the Prevention and Eradication of Commercial Sexual Exploitation of Children, the Prevention and Suppression of Prostitution Act of 1996 of Thailand already ensured support in the prevention stage (Cotter 38). In 1999, under the influence of the National Commission on Women's Affairs, Thailand became the first country in Southeast Asia to pass laws to impose graver penalties on customers than on sellers for engaging in commercial sex with underage kids ("Prevention," par. 5). The Thai government also passed the Anti-Human Trafficking Bill in 2007, which not only ensures that victims of trafficking get substantial protection, but also punishes violators in more severe terms ("Thai Government," par. 15).

Following the lead of the Thai government, other Southeast Asian countries soon carried out a series of other laws and acts: the Penal Code from Laos and Vietnam prohibits any kind of abduction and trade in person; the Law on the Suppression of the Kidnapping, Trafficking and Exploitation of Human Beings is adopted by the Cambodian government to address the issue; Myanmar also has similar laws regarding sex trafficking and commercial sex activities: the Child Law (1993), the Suppression of Prostitution Act (1949) and the Penal Code (Cotter 38 - 45).

C. Actual Combative Efforts in Southeast Asia

Despite the fact that the governments in Southeast Asia have signed the declaration against human trafficking, not much real progress has been made, because of the complicated political status quo in that region, such as ethnic conflicts, coups, civil wars.These factors continue to pose threats to the effectiveness of anti-trafficking measures. In the TIP Report, countries are placed into different tiers according to their dedication to fight human trafficking. In Tier 1 are the countries with the anti-trafficking participation above the minimum standards, while in Tier 3 are countries with the worst anti-trafficking performance. For the five major countries with a human trafficking issue, Cambodia is in Tier 2, Laos, Thailand and Malaysia in Tier 2 Watchlist, Myanmar is in Tier 3 ("2016 Trafficking" 58), meaning none of them meets the minimum requirements to combat human trafficking.

In the newest TIP Report, although the Myanmar government, for example, continued to investigate cross-border sex (and labor) trafficking crimes and collaborated with international partners to identify and demobilize children who were forcibly recruited in to the military, its overall victim identification and protection remained weak. Lack of survivor protection services left victims highly vulnerable to re-trafficking ("2016 Trafficking" 113). In addition, all five countries have shown different levels of government corruption, such as the prosecution department not holding military and civilian officials who engaged in the crimes accountable ("2016 Trafficking" 113, 119, 238, 364). But fortunately, the Thai government has already improved the corrupt system by contributing increasing effort to convict government officials who are complicit in trafficking crimes (364). More details about government corruption will be discussed in the next section as one of the limitations that social media measures would not be able to address.

IV. Practical Recommendations Regarding Social Media

Although ASEAN countries have signed the Declaration Against Trafficking in Persons, Particularly Women and Children, no significant progress has been made on implementing the Declaration in Southeast Asia. This is because of the complexity of trans-regional cooperation, a challenge posed by the complicated political status quo in Southeast Asia. This is why innovative ways are needed to tackle the issue: we should turn to technology. In the 21st century, technology has been playing an increasingly important role in all fields. Not surprisingly, traffickers are taking advantage of social media platforms and online classifieds to engage in sex trafficking, especially for recruitment, advertisement and communication purposes (Latonero 8**)**. Traffickers have more online targets now than ever — especially young women and men. Since they are the mainstream users of social media, they often become the targets of the traffickers. Not only sex trafficking, but also human trafficking, have become increasingly intertwined with technology and therefore more diffuse and adaptive than its traditional face-to-face form. Donna Hughes, a well-known American researcher on women and children trafficking observes, "the sexual exploitation of women and children is a global human rights crisis that is being escalated by the use of new technologies"(Latonero 14**)**.

Let us now consider how technology can combat sex trafficking. Trafficking's growing dependence on technology casts new light on how we can deal with sex trafficking in a more effective way, that is with social media and other online platforms. In this section

of the paper, we will discuss kinds of social media and how they can be utilized to combat sex trafficking in Southeast Asia. However, it is also important to notice that social media as public platforms still have some limitations and cannot address every aspect of the issue.It therefore is just one possible tool, not an absolute solution, to fight sex trafficking.

D. What It Means by Social Media

Social media, by definition, are electronic communicating tools through which people can use online platforms to share and exchange ideas, information and communicate messages ("Full Definition of Social Media," par. 1). In this paper, social media include online social networking sites such as Facebook, Twitter and MySpace, online platforms like BBS, chatrooms and forums, and apps on phones.

E. The Three Merits of Social Media

Some people may wonder what makes social media effective tools in the fight. I have considered three characteristics of social media that can help the current struggle. The first merit is high public participation. The public's power can be big enough to make a difference. As of 2015, ASEAN countries had 252.4 million active Internet users, which is about 40% of Southeast Asia's total population ("Southeast Digital," par. 5). In some countries like Vietnam, Internet users are as high as 52% of its total population ("Internet Users by Country"). This big user group is a merit: with such high internet penetration, even if each Internet user devotes only a little effort, it could make a huge difference.

The second characteristic, which goes along with public participation quite well, is the easy accessibility of social networking sites. We do not need fancy technology that requires huge data calculation; we can easily access social media as long as we have phones, which probably many people in Southeast Asia already have.

Last but not least is its high speed. Through a click of a finger tip, messages can be sent to cousins who live thousands of miles apart within a second. In this way, whenever we detect suspect traffickers or encounter trafficking victims, we can report to the authorities or to NGOs in an instant — which saves a lot more time than traditional reporting terms.

With these three merits, social media can be a powerful tool to mobilize the power of the public and make full use out of it to combat sex trafficking.

The rest of this section will demonstrate how and where social media can be used in the three stages of the 3P paradigm along with examples and cases. The targets of these recommendation include victims and survivors of trafficking, customers, bystanders and governments. The recommendations have certain levels of overlap in measures and in audiences. But it is also noteworthy that the effort requires multilateral, not unilateral cooperation. With different weight put on different audiences in the three stages, this section attempts to give a well-rounded perspective on practical suggestions to fight sex trafficking in Southeast Asia.

F. Structural Obstacles

There are some structural obstacles in practice that might hamper the effectiveness of social media in this fight: poverty and illiteracy-led inaccessibility and corruption.

a. Inaccessibility:

Speaking of social media in Southeast Asia, many people would doubt the scope of this method: how many people would actually benefit from social media? Many of them are just too poor to own a smart phone. Even if they have access to electronic devices, how many of them are educated enough to be able to use smart phones and computers?

Figures 6 and 7, the global map of adult and youth literacy rate of 2012 from UNESCO Institute for Statistics, show that Southeast Asia was actually doing rather well in literacy even in 2012. While Laos and Cambodia have an 80% - 89% literacy rate, all other countries are in the 90% - 99% category, which means that most people in Southeast Asia have received some basic education. Giving them electronic device training, such as providing them with user's guides or instructions, would not be a big problem. Even if they have difficulties using apps and social networking that require complex technology unfamiliar to them, simple operations can be devised, which will be discussed later in the recommendation section.

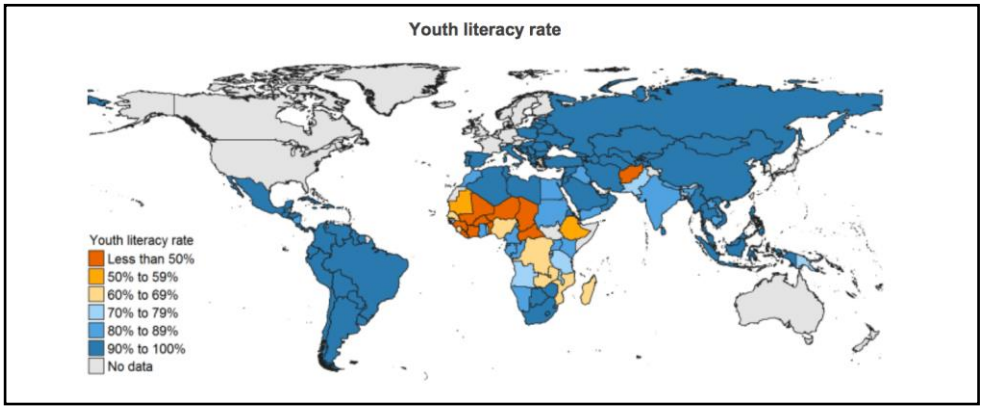


Figure 6: *2012 World Map of Adult Literacy Rate*
(Source: UNESCO Institute for Statistics)

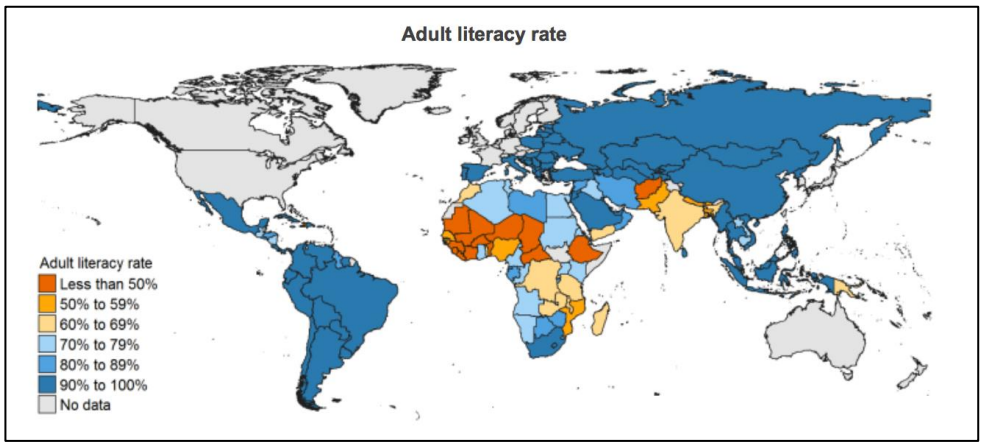


Figure 7: *2012 World Map of Youth Literacy Rate*
(Source: UNESCO Institute for Statistics)

How about poverty? It is true that people from rural areas are typically poor and are not educated, but they are often trafficked to sex sector in urban areas, where there are plenty of accesses to electronic device. Though the victims themselves might be prevented from smartphones and computers, people living there can give a hand at this very moment. What's more, as previously mentioned, a relatively simpler and smaller device can be created that would be very easy to carry but will not be suspicious.

b. Corruption:

The other limitation is the corrupt governments that obstruct progress. Criminal groups bribe authorities with wealth and sexual favors, and government officials involved in sex trafficking provide shelter for the criminal groups in return. This loop continue to be a big threat that impedes progress. As early as 2007, researcher Sheldon X. Zhang and Samuel L. Pineda verified the significant correlation between the general levels of corruption and the extent of trafficking within a country (qtd. in "The Role of" 12). Zhang and Pineda later concluded in their research that "corruption is probably the most important factor in explaining human trafficking" and that "[c]ountries that make the least effort to fight human trafficking also tend to be those with high levels of official corruption" (qtd. in "The Role of" 12). While the extent of corruption in a country is determined by the country's tier placement (Tier 1, 2, 2-watchlist, 3, and special case) in the TIP Report (Trafficking in Persons Report), their research shows a statistically significant negative relation between the ranking and the levels of trafficking in persons: the lower a country's ranking is, the higher extent of trafficking it has.

Discussion of sex trafficking in Southeast Asia oftenrefers to Thailand, Vietnam, Laos, Myanmar (Burma), Malaysia and Cambodia. Their rankings in the 2016 TIP report are as follows: Cambodia is in Tier 2, Laos, Thailand and Malaysia in Tier 2 Watchlist, Myanmar is in Tier 3 ("2016 Trafficking" 58). None of them is meeting the minimum standard for the elimination of trafficking — they are all having some kind of trouble within the government. Specifically, in the 2015 CPI Report (Corruption Perceptions Index Report), none but Malaysia has a score higher than 40 out of 100, indicating a severe level of government corruption. Among them, Cambodia has the lowest score: only 21 ("Corruption Perceptions Index," par. 24). The authorities give warnings to brothel owners in exchange for sexual favors and bribes; many high level officials abuse their power by letting the traffickers and exploiters go and getting only the victims; knowingly, the governments do not prosecute officials complicit in trafficking. In addition to these practices, respondents in the Zhang and Penida research suggested that border patrols, law enforcement agencies, and the police are the three most vulnerable positions to corruption (qtd. in "The Role of" 12). For example, according to the TIP Report 2016, some government officials in Thailand are directly profiting from bribes and direct engagement in blackmailing migrants and their sale to the traffickers ("2016 Trafficking" 366); high-level officials in Myanmar even directly tell the victims of trafficking not to ask for the government's help ("2016 Trafficking" 116). Such corrupt practices cultivate impunity for the traffickers, making it increasingly harder for the masses to guard against trafficking. It is therefore necessary to mobilize the public's support for the combat and multilateral monitoring system and cooperation. Corrupt governments will hardly help any of the victims.

G. Practical Recommendation

a. Prevention:

1. The most important task in the prevention stage is to raise public awareness against sex trafficking. If people are conscious of the potential risk of sex trafficking and are informed of how to avoid being trafficked and how to identify sex traffickers, there would probably be a substantial decrease in victims. On the other hand, if customers of sex tourism are aware of the harm brought by their action, they would stop facilitating this crime by not engaging in it. Therefore, awareness raising campaign should be the main focus in this stage.

Social networking sites such as Facebook and Twitter can be perfect platforms for anti-trafficking NGOs, governments, and even individual trafficking fighters to share ideas and projects against trafficking. Facebook, by far the most commonly known and the most widely used social networking platform, has over one billion users around the world. Because of its huge number of users, Facebook has gradually become the hub of "socially conscious networking" ("Digital Activism," par. 9), promoting all kinds of social awareness campaigns. For instance, anti-trafficking NGOs can create Facebook groups and pages to share real time projects and developments and to post related stories and pictures, whereby they can spread anti-trafficking knowledge, promote anti-trafficking efforts bit by bit and expand their audience. DNA Foundation, an NGO devoted to combatting child sex slavery, for example, has a Facebook page with more than 110,000 "likes," and other organizations like ECPAT UK has more than 29,000 "likes" (par. 9). If people are interested in knowing more, they can easily "join the group" on Facebook page. These NGOs use Facebook not only as a platform to share knowledge, projects and developments with the general public but also an effective tool to call for action and communication with organizations having similar initiatives (par. 9).

Twitter, too, could be thought of as a center for initiatives and struggles against human trafficking, hosting numerous influential discussions around human trafficking issues. Some noted organizations using Twitter against human trafficking are receiving close attention: Not For Sale has about 62,400 followers, the Polaris Project has approximately 40,000 followers, the CNN Freedom Project has around 15,600 followers. To awake public awareness, all these accounts need to do is tweet relevant trafficking information like the ones shown in the picture down below, with links that can redirect followers to websites where they can take action or learn more about trafficking (see figures 8 & 9).

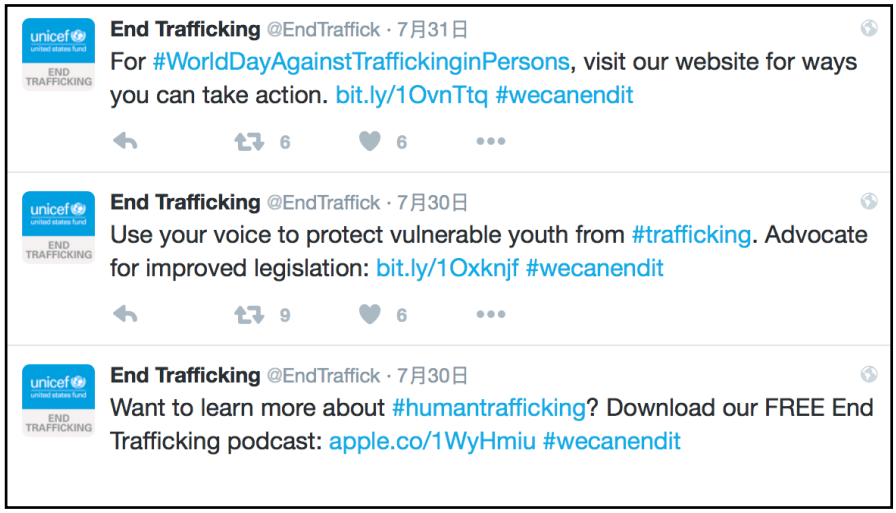


Figure 8: *Snapshot of Twitter - End Trafficking*
(Source: Twitter)

Figure 9: *Snapshot of Twitter - Polaris Project*
(Source: Twitter)

Even survivors can play a role in raising social awareness. Somaly Mam is a good demonstration of how a survivor can promote social awareness. She is an anti-trafficking activist and a survivor of sex trafficking from Cambodia, and is now sharing her stories and her work through the links she tweets. Her regular tweets include trafficking stories, events, and possible opportunities for action. For example, she alerted her followers in consecutive tweets in 2012 that her organization's centers in Cambodia had received three trafficked women. By sharing efforts to combat sex trafficking, Mam illustrates through her tweets what she has been doing in the struggle and along the way mobilizes potential support around volunteer work ("Digital Activism," par. 12). Despite work opportunities, though, survivors could do more to prevent sex trafficking in the very first place. If they are willing, survivors can share their stories of trafficking through social networking sites, forums, and even chatrooms. Through story telling, more people would acquire firsthand evidence as to how and where the victims are trafficked, being cautioned of potential risks of sex trafficking, and may even apply this knowledge to identify victims and traffickers. They could even hold realtime chats online — where the public can actually question survivors —

just like online interviews. In this way, people can get to know more about a specific aspect. However, participants should be careful when asking a survivor to do such things.It is not always a good idea for survivors to discuss their traumas again, as some of them might suffer mental loss a second time during story sharing.

In sum, social networking platforms would be a noteworthy venue for raising public awareness against sex trafficking and cautioning people, especially young men and women of the potential traps that traffickers may pose. Through mainstream media, we can share ideas and knowledge about sex trafficking, mobilize support for anti-trafficking effort, and promote volunteer work (to aid against trafficking).

2. Some might doubt the effectiveness of social media as a platform to raise awareness since some poor families have no access to high technology like cell phones and computers. Indeed, such methods can only reach people who can afford cell phones and computers — but that is already addressing half of the population. The other half could be addressed by a second suggestion: the government should outlaw the use of social media as a tool of recruitment, advertisement, and communication during sex trafficking. Websites such as Craigslist and MySpace have online classified advertisements labeling "adult service" or "erotic service," through which traffickers can lure innocent juveniles or even adults into the life-altering traps (Latonero 21). The best way to prevent sex trafficking is to block every possible way for the traffickers to take advantage of the vulnerable in the very first place.

b. Prosecution:

The key to prosecution is crime detection. As long as we detect the crime, we have made our first step towards tackling it; after detection, we should do something about it, like report suspects and prosecute criminals. However, governments and authorities alone cannot effectively address this task alone.This is where the public's power should be called for. Collaboration between multiple stake-holders is essential in detecting crimes. In this stage not only can the authorities contribute their own effort, so can customers and even bystanders. Given that the victims have already fallen into the hands of the traffickers by this stage, the following recommendations will focus mainly on customers, NGOs and government agencies, as well as bystanders.

1. Data-mining techniques, including term-frequency analysis (Latonero 27) and facial recognition (Latonero 30), would have an important place in detecting potential trafficking cases. Facial recognition enables the police to sift through billions of online classified advertisements to track down subjects, such as reported victims. Because traffickers would usually upload the victims' pictures online for advertisement, with such technology, it can be determined whether or not the "reported victims" truly fell victim and where they might be. However, facial recognition technology needs to be developed into a more mature state if it is ever put to use, since it is still a challenge to rely purely on a picture prototype to match faces. Even a minor change in poses or angles will make the currently sensitive recognition technology miss the target (Latonero 30). But what if the advertisements do not contain pictures? Are crimes still be detected? The short answer is yes, but with a different technology. Term-frequency analysis reveals the most commonly used terms in advertisements; then by searching these key words through the search function in the database of any social media platform, it is easy to find online "adult service"

advertisements (Latonero 29). For example, in a research study featuring Twitter as a potential platform to detect trafficking, CCLP (Center for Children's Law and Policy) and ISI (Information Sciences Institute) researchers initiated a search of public tweets through the searching function. The study centered on the key word "escort," and they collected data for a one-week period.The team collected 315 posts containing "escort" services. Next they used term-frequency analysis to analyze which terms these posts, besides "escort," contained the most. It turned out that terms describing the victims' nationality appear the most, like "Indian," "black," or "Asian." A more in-depth analysis suggested that the posts described female victims' physical characteristics in details, such as "young" and "tiny" (Latonero 27). Such methods not only help the police detect potential crimes of sex trafficking but also help narrow down the pool of suspects and victims. However, these two technological methods alone cannot guarantee detecting sex trafficking with certainty. A nuanced difference in pictures could make a huge impact, and the different meaning of "escort" would make the detection stray off course. Therefore, I suggest when employing these two methods, technology should be combined with human work. Equip manual face recognizers with the facial recognition machine; ask the linguists for help before term searching.

2. The above two measures require the aid of high technology. The following paragraphs will introduce two accessible measures for the general public to detect sex trafficking as well as to report the crimes to the authorities and to help the victims. One is crowdsourcing, and the other is mapping.

Although "crowdsourcing" seems to be very technical, it is actually a simple concept. The term is defined as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" (qtd. in "Digital Activism," par. 6). In terms of social media, we can simply put it as an act of mobilizing Internet users to finish a task (in this case sex trafficking detection) together. In crowdsourcing, anti-trafficking organizations, activists and the fellow individuals may form an online community against sex trafficking. Whenever any member posts a help-seeking feed, other members in the social net will give a hand until they successfully pull the victims, the to-be-victims, and the survivors out of the plight. A famous example of using crowdsourcing to tackle trafficking issue happened in 2010. Dan Reetz suspected his former student, a young Russian women and her friend, were about to be lured into a human trafficking operation when he heard that the "exchange program" insisted on changing the meet-up place from Washington D.C. to New York. The two girls had paid $3,000 for a "promising" job as hostesses in D.C. and would not listen to Reetz. Worried, Reetz then posted "Help me help my friend in D.C." through the MetaFilter discussion, which is still available on line, trying to get people to convince his students not to fall into the trap. Within minutes after posting, members on MetaFilter offered valuable information. Luckily, a member called Kathrine Gutierrez Hinds volunteered to meet the two Russian girls at a New York bus depot, and eventually convinced the two girls not to meet with the "exchange program." A trafficking groups are gradually applying crowdsourcing tactics to their missions, and numerous cases like Reetz's suggest a huge potential for crowdsourcing to become a tool to fight sex trafficking (pars. 6 - 8).

Mapping, also called flagging, is to label a location with a symbol, usually a flag. When identifying a potential hub for sex trafficking such as brothels, the public, especially

the customers, could bring their advantages into play. Actually, a Southeast Asia country once had the experience of using mapping to combat crimes. In 2012, Operation Endeavor, a multilateral cooperation between US, UK and Filipino law enforcement agencies, NGOs and corporation effort employed data mapping to identify regions in the Philippines through which child sex abuse material was transported. With this technology, Operation Endeavor discovered and dismantled crime groups engaged in child abuse in 11 Filipino sites. Though this case does not technically fight sex trafficking, it does provide valuable insight into how mapping could be used to identify possible sex trafficking hubs (Cook and Heinl, par. 9).

A more direct case is the first anti-trafficking mobile phone application. In 2013, just a year after Operation Endeavor, an app was developed by RedLight Traffic to make Anti-Slavery Day in London come true. The app cooperates with the Polaris Project, equipping users with potential trafficking indicators so that users can use red flags to identify victims. To ensure some uneducated users can correctly use the software, it has a 20-minute video teaching users how to operate the app on the smart phone. Its two most obvious merits are that it enables users to anonymously reports suspected cases to the local police's, and it provides a sharing platform where users can establish an online community network against trafficking. In this way, not only the reporters' safety is guaranteed, but the momentum against sex trafficking can be stored through the online social network, since the local community can form its own online community network to note the effort each member has made (Cook and Heinl, par. 10). To avoid a situation where a corrupt government, which is facilitating sex trafficking, does nothing upon the signals, the anonymous reports should be delivered to local NGOs that have the same cause, where the victims can actually be rescued.

If the public cannot identify victims, there is another way for victims to save themselves, as long as they have a signal transmitter as small as a button. With a signal transmitter in hand, whenever they feel they are in trouble, they just press the button to emit the radio (or whatever signals it has) on the transmitter, so that they can inform the rescuers in their locations. Again, the signals would be anonymous, labeled as serial numbers, and the site would be off high security to protect the victims. Also very similarly, the signal would be directed to NGOs and to anti-trafficking activists to ensure rescue. The operation is very easy to learn, and usually such devices are affordable for the general public. There is then no need to worry about whether victims can afford smart phones or whether they are literate. However, this measure requires a considerable amount of funding to do all sorts of purchasing, equipping and publicizing, which in Southeast Asia is a little bit difficult. But still, this is one perspective to look at it and is worth trying.

The above methods center around high technology: data analysis, crowdsourcing, and mapping. But such initiatives are sustainable only if multiple stakeholders play their parts. Therefore, interregional, inter-party and inter-class collaboration is essential in this stage.

c. Protection:

Finally, we have come to the final stage of the 3P strategy: protection. In this stage, the victims, who should now be called survivors, have been rescued. As mentioned in previous sections, after the girls are tricked by traffickers, the common destination for them are brothels. They are sold to customers for $30 per night, for example, but they barely

receive a penny. In addition, they are often beaten, raped and even tortured by their pimps. Physically and mentally abused, it is very common for these survivors to develop some kind of illness. And often, even if they are still healthy after they are trafficked, many of them are too poor to sustain their basic living, or they are not educated enough to find themselves a job. This is why they need they help from the public, the NGOs, and thousands of anti-trafficking activists. It is to the people who can to help the survivors rebuild lives of dignity, stability, and economic autonomy. The questions remain: where should the survivors go? how to ensure they do not fall victim again? even if they are rescued, how can they be provided with a different (and brighter) life?

1. First things first.They should be ensure the basics of living --shelters, food and clothes. To realize this resource distribution, initiators need an online platform and a local resource-distribution center. Again, this is where crowdsourcing could play to its potential: on the online platform, government agencies and NGOs could mobilize every resource available by first gathering available resources from a large group of members, then sort them out by different classifications (shelters, food, clothes, job opportunities, etc.), so that when the survivors come for help, workers and volunteers at the center can easily distribute the required resources. The survivors only need to know where the nearest distribution center is located to ask for help. There is no need to know complex technology, and no need to pay huge amounts of money. Despite guaranteeing their basic survival, survivors should be empowered with economic ability. Again, crowdsourcing could find suitable jobs for the survivors with just a click of a finger in the distribution center. Such platforms are not limited to the locals; they are connected to each other on the platform, so that when one center falls short of certain resources, the survivors can be directed to other centers.

A workable example of such a platform is Twitter. The A21 Campaign, a non-profit organization aimed at ending human trafficking, launched an operation in 2007 to provide shelter and tradition homes for survivors of human trafficking in Greece, Ukraine, and Bulgaria. It was Twitter which ensured the success of this campaign: it provided a way for the A21 Campaign to regularly connect with and contact a large audience which was expanding with time (“Digital Activism,” par. 12). Although this campaign’s operating region is in Europe, I believe Southeast Asia can borrow this basic operating mode, with proper adaptations to local environments. Nomi Network, also a non-profit organization, has worked similarly in Asia. It provides entrepreneurship, leader ship, and technical skills to trafficking survivors, training them to become a fitter worker in the modern market (“Program Model,” par. 1). In 2009, for example, Nomi Network began to work with Cambodian enterprises and training centers to prepare sex trafficking survivors as well as potential victims for transitional jobs. Nomi’s aid successfully secureed 23 women’s jobs, and by 2011 it created 80 more jobs for the survivors (par. 6).

2. What could help the survivors say goodbye to the memory of their sufferings? Consolation, comfort, and proper treatment if necessary. Traumas associated with human trafficking can be detrimental and enduring. But NGOs and governments can organize and operate online chat rooms and forums for the survivors to go to. By pouring their sufferings and grief out in the chat rooms, constantly giving support to each otherin survivor forums, survivors could gradually let go of the miserable memories — or at least learn to cope with them. However, the members of chat rooms and forums had better be exclusively sex trafficking survivors, with one or two psychiatrists and hosts. This is because in Southeast Asia, as previously mentioned, a lot of people still hold patriarchal values and heavy biases

against women, especially women who were sexually abused by many people. These unfortunate women normally are burdened with the stigma of being "dirty". To truly protect the survivors' dignity and mental health from being hurt, disrespectful people should be excluded from the discussion. Also, a psychiatrist at the site can guide survivors to adopt a healthy mindset and mood, and teach them scientific ways to help relieve the sense of sin and pressure, which many survivors have. Through forums and chat rooms, volunteer workers and experts can also caution survivors about ways to detect sex trafficking traps, prepare them to reenter society, and encourage them not to lose faith in life. As for law enforcement agencies, it is their responsibility to regulate online social platforms where some graphic pictures would commodify women as goods by labeling them with bar codes (as figure 10 shows). Such pictures, along with other similar comments would harm the survivors twice. Thus, it is desirable if policy makers in Southeast Asia outlaw the use of such photos and regulate similar remarks online. Regulators can employ labeling system to achieve regulation on the Internet effectively by labeling suspicious pictures and remarks.

Figure 10: *A Commodified Woman*
(Source: Human Trafficking Center)

V. Conclusion

As one of the most lucrative crimes on earth, sex trafficking poses a great threat to human rights, whether in Southeast Asia or in the whole world, and it is encouraging to see so many efforts working together to combat sex trafficking from three perspectives: prevention of trafficking, prosecution of the traffickers, and protection of the victims (3P paradigm). Having signed the numerous documents on protecting human rights such as UN Declaration of Human Rights, Southeast Asian countries are also obligated to fulfill their responsibilities. However, given the current corrupt governments in place, it is hard to rely solely on the authorities to crack down on sex traffickers. It is time for the public to act. Together on social platforms, with technological advancements including social networking sites (Facebook and Twitter), facial recognition, term-frequency analysis, mapping, crowdsourcing, online chat rooms, forums, and signal transmitter, the global civil society, and society in Southeast Asia in particular, can fight sex trafficking in all three stages.

Moreover, acknowledging such structural limitations as poverty, literacy and corruption is important for the actual implementation of the recommendations in this paper. Literacy and poverty are generally not considered problems since simple and affordable methods can be taught to rural residents in Southeast Asia. However, despite minor improvement in practice in some countries, corruption in general remains an essential issue to be dealt with in this fight, and governments still need to work on logistics to ensure the efficacy of measures taken.

As Kevin Bales has put it: “Slavery, without us even noticing, has ended up standing on the precipice of its own extinction, waiting for us to give it a big boot and knock it over. And get rid of it. And it can be done” (Bales). Although such structural obstacles as government corruption and inaccessibility to social media might temporarily hinder our way towards justice, eventually, evil will give way. With the use of social media as weapons, global civil society can greatly diminish sex trafficking.

Bibliography

"2016 Trafficking in Persons Report." *United States Department of State*. Washington D.C., 2016, pp. 4, 58, 113, 116, 119, 238, 364.

"3Ps: Prosecution, Protection, and Prevention." *United States Department of State*. net. July 09, 2016. <http://www.state.gov/j/tip/3p/index.htm>.

"About the Convention." *Council of Europe.* net. Aug. 04, 2016. <http://www.coe.int/en/web/anti-human-trafficking/about-the-convention>.

Annan, Kofi. A. "United Nations Convention Against Transnational Organized Crime and the Protocols Thereto - Foreword." *United Nations Office on Drugs and Crimes.* Vienna, 2004, pp. iv.

"ASEAN Declaration Against Trafficking in Persons Particularly Women and Children." *No Trafficking.Organization.* Nov. 29, 2004. net. Aug. 03, 2016. <http://www.no-trafficking.org/reports_docs/lao/laws/ASEAN-Declaration-29-November-2004-ENG.pdf>

Bales, Kevin. "How To Combat Modern Slavery". 2010. Speech.

Cook, Alistair D. B. and Caitriona H. Heinl. "Human Trafficking in Asia Going Online." *East Asia Forum.* May 03, 2014. net. Aug. 02, 2016. <http://www.eastasiaforum.org/2014/05/03/human-trafficking-in-asia-going-online/>.

Cotter, Kelly M. "Combating Child Sex Tourism in Southeast Asia." 2009, vol 33:3, pp. 38 - 45, 497.

"Corruption Perceptions Index." *Transparency International Organization.* net. Aug. 03, 2016. <http://www.transparency.org/cpi2015>.

Couch, Robbie. "Human Trafficking Is Still Globe's Fastest-Growing Crime Despite Increased Awareness." *The Huffington Post*. Jan 07, 2015. net. July 02, 2016. <http://www.huffingtonpost.com/2015/01/07/human-trafficking-increasing_n_6425864.html>.

"Digital Activism in Anti-Trafficking Efforts." *University of Southern California.* net. July 31, 2016. <http://technologyandtrafficking.usc.edu/digital-activism-in-anti-trafficking-efforts/>.

"Full Definition of Social Media." *Merriam-Webster Dictionary.com.* net. July 31, 2016. <http://www.merriam-webster.com/dictionary/social%20media>.

"Global Report on Trafficking in Persons" *United Nations Office on Drugs and Crimes.* Vienna, 2014, pp. 34, 43.

"Global Safety Net." *The Polaris Project.* net. July 04, 2016. <http://polarisproject.org/initiatives/global-safety-net>.

Harden, Nathan. "Eight Facts You Didn't Know about Child Sex Trafficking." *The Huffington Post.* Nov. 11, 2013. net. July 02, 2016. <http://www.huffingtonpost.com/nathan-harden/eight-facts-you-didnt-know_b_4221632.html>.

"Human Trafficking." *United Nations Office on Drugs and Crime.* net. July 02, 2016. <https://www.unodc.org/unodc/en/human-trafficking/what-is-human-trafficking.html>.

"ILO Global Estimate of Forced Labour - Results and Methodology." *International Labour Office*, Geneva, 2012, pp. 13, 77.

"Internet Users by Country." *Internet Live Stats.com.* July 01, 2016. net. Aug. 02, 2016. <http://www.internetlivestats.com/internet-users-by-country/>.

Kabance, Karie. "The Globalization of Sex Trafficking." *International Affairs: Directed Research Project.* 2014, pp. 12, 18.

Latonero, Mark. "Human Trafficking Online The Role of Social Networking Sites and Online Classifieds." *Annenberg School of Communication & Journalism, University of Southern California*. California, 2011, pp. 8, 14, 21, 27, 29, 30.

"Prevention." *Human Trafficking Organization.* net. July 05, 2016. <http://www.humantrafficking.org/combat_trafficking/prevention>.

"Profits and Poverty: The Economics of Forced Labor." *International Labor Organization.* Geneva, 2014, pp. 15.

"Program Model." *NomiNetwork.org.* net. Aug. 02, 2016. <http://nominetwork.org/impact/>.

"Sex Trafficking." *The Polaris Project.* net. July 02, 2016. <http://polarisproject.org/sex-trafficking>.

"Southeast Asia Digital, Social and Mobile 2015." *ASEAN up.com*. March 28, 2016. net. Aug. 02, 2016. <http://aseanup.com/southeast-asia-digital-social-mobile-2015/>.

"Success." *Polaris Project.* net. July 08, 2016. <http://polarisproject.org/successes>.

"Survivor Support." *The Polaris Project.* net. July 04, 2016. <http://polarisproject.org/initiatives/survivor-support>.

"T Visa." *Immigration.com.* March 10, 2009. net. Aug. 04, 2016. <http://www.immigration.com/visa/t-visa/t-visa>.

"Thailand." *Human Trafficking Organization.* net. July 03, 2016. <http://www.humantrafficking.org/countries/thailand>.

"Thai Government and International Organizations Pledge Cooperation to Provide Assistance to Victims." *Human Trafficking Organization.* June 04, 2007. net. July 03, 2016. <http://www.humantrafficking.org/updates/653>.

"The Global Initiative to Fight Human Trafficking (UN.GIFT)." *United Nations Foundation.* net. July 04, 2016. <http://www.unfoundation.org/how-to-help/donate/ungift.html?referrer=https://www.google.com/>.

"The Role of Corruption in Trafficking in Persons." *United Nations Office on Drugs and Crimes.* Vienna, 2011, pp. 12.

"Universal Declaration of Human Rights." United Nations. net. July 03, 2016. <http://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf>.

"UN.GIFT." *United Nations Global Initiative to Fight Human Trafficking Organization*. net. July 03, 2016. <http://www.ungift.org/knowledgehub/en/about/index.html>.

"UN.GIFT - United Nations Global Initiative to Fight Human Trafficking." *United Nations Office on Drugs and Crimes.* net. July 04, 2016. <https://www.unodc.org/lpo-brazil/en/trafico-de-pessoas/ungift.html>.

"Wages in Asia and the Pacific: Dynamic but Uneven Progress." *ILO Regional Office for Asia and the Pacific.* Bangkok, 2014, pp. 2.

"What Is Trafficking in Persons." *Office to Monitor and Combat Trafficking in Persons, United States Department of State.* Washington D.C., 2012, pp. 1 - 2.

The Market Efficiency of "Smart Money" During the Tech Bubble

Kevin Li

Author background: Kevin Li grew up in the United States and currently attends Naperville North High School, located in Naperville, Illinois. His Pioneer seminar topic was in the field of economics and titled "Stock Market Crashes."

Introduction

It is remembered today as an anomaly; in retrospect, market participants and professional investors look back with confusion and can only answer for their actions with unsatisfying hindsight. Few, if any, can say they saw the events that transpired coming.

It was a time of jubilance and optimism in the wake of exciting technological development and opportunities for a new era. *Newsweek* said that the "digital revolution would create a zillion dollar industry."[8] Prices of securities, especially those affiliated with the technology and internet sector, were inflated to unprecedented levels. But when the music stopped, losses were equally astonishing. The sentiment pulled a complete reversal. Many burgeoning startups shut down operations, and people came to the realization that their initial optimism for the tech industry was far too exuberant.[9]

These types of price movements do not happen without the appearance of severe scrutiny and analysis, and deservedly so. Bubbles in any form and to any extent challenge the theories that underlie accepted principles of financial markets and create considerable disappointment for all parties involved. Animosity is commonly stirred up between conflicting interest groups as fingers are pointed and blame placed. As retail investors take massive losses, they cannot help question how Wall Street has fared, and/or if Wall Street performed some form of morally reprehensible "insider scheme" that should be punished.

Nonetheless, a typically accepted idea that has taken hold in the minds of market participants and the general public is that industry professionals, the ones with access to larger amounts of capital, better and faster information, experience, and raw intelligence are the ones that generally stay out of "market folly." According to the efficient market hypothesis, professionals are supposed to be the ones that mediate irrational noise trader perception on both sides of the spectrum, taking the contrarian view of both bullish and bearish action to make sure that securities are fairly priced and free of irrational emotion. The hypothesis implies that institutional investors, therefore, do not suffer anywhere close to the full extent of rapid price movements in the financial markets presumably fueled by emotion.

To be fair, the hypothesis is fairly reasonable. To say that professional investors behave like noise traders and retail investors would be implying that their academic credentials, their college degrees, and their years of experience are all for naught, and do not count in terms of their ability to maneuver markets at a level above the general public. To take the implication that far would be a misleading statement that forgoes common sense. One must be rational when making broad statements that judge such a significant section of market participants.

However, evidence suggests that financial institutions did not fare much better or act much faster than the general public during the Dotcom mania. The bubble has been

shown to have been a crazed time of Greenspan's "irrational exuberance" for all parties involved in the markets; the professionals actually did not make sure that securities were valued according to their intrinsic financial value.[4] They made the same mistakes as retail traders and showed no significant signs of foresight into what was about to occur. Losses were spread between them. They exhibited the same patterns of herd mentality as those with far less experience.[7]

Institutional portfolio managers and retail traders show little difference in their reaction to and performance during the Dotcom bubble. The contrarian tendencies institutions were meant to exhibit by remaining emotionless and fairly pricing securities, regardless of the irrational activity by noise traders, was scarcely found in the entirety of the bubble's cycle. It can be concluded, therefore, that the efficient market theory became null and void during the bubble.

Literature Review

To delve into a topic as complicated and controversial as the Dotcom bubble requires a significant amount of background knowledge and understanding of various opinions. There is no simple justification or explanation for a bubble; by nature it is a culmination of multiple situations or catalysts. Bubbles are complex and draw various explanations.

Nonetheless, there are quite a few academic papers that do an excellent job of compiling a significant amount of varying opinions and conclusions and giving a detailed overview of the actions and thought processes of various parties. One such article is written by Goodnight and Green (2010). It summarizes the events that led to the dot-com bubble and its inevitable bursting. The paper goes through a chronological retelling of the various phases which made up the crisis and gives various citations and reasons as to why each phase happened, along with commentary from prominent media sources that had significant influence on the public at the time. There is no obviously discernible conclusion that can be made from reading the paper. However, various theses and points of view are presented.

One of the main ideas was the implication of blame that was attributed to various parties. Throughout a considerable portion of Goodnight and Green's paper, there are pieces of editorials and newspaper commentaries that suggest that retail traders felt trauma, anger, and confusion after the bubble popped. Of course, this is not a surprise, but an idea formed in trying to determine who really was at fault and if retail investors were justified in their angst.

A paper by Griffin, Harris, Shu, and Topaloglu (2011) identified insightful research. It gives a general commentary on the tech bubble overall but focuses on relevant parties, giving an overall verdict on who should claim the most responsibility for it. The significance of individual and institutional investors is examined and their relative importance is considered. They analyze order flow from the two parties, and they find that institutional investors have a larger purchasing volume in the lead-up to the bubble than retail investors. This contradicts certain conditions for models explaining the cause of the bubble made by other academics, especially those claiming that bubbles are due more to the individual investor than the professionals. Conclusions are made by Griffin et al. that in the lead-up to the tech bubble, institutional investors actually do trade independently of future fundamentals of the relevant tech companies, that they engage in riskier stocks and companies, and that they pull out before retail investors when the collapse begins. It culminates to state that institutional investors are the main cause for the tech bubble crash.

However, there is still skepticism that when analyzing a paper that condemns institutions for not following what their specified role should be - an efficient moderator. It

could very well be that companies had far higher growth projections than underlying performance in the bubble and simply had to deal with far lower growth projections and performance in the "pop." That notion would have still have made sense and justified the actions of institutions; their actions performed before retail traders would have been simply a sign of faster and more efficient use of information and pricing.

Wheale and Amin's paper examining investor behavior during the tech bubble (2003) provides strong evidence to support the notion that the efficient market theory could not feasibly be applied to the rancor that surrounded the volatile price action of the Dotcom bubble. They bring forth multiple lessons that investment managers should learn from their findings. One of their main conclusions is that only *some* of the relevant metrics used to value tech securities were considered in the lead-up to the Dotcom bubble, yet *all* metrics were considered in the selloff period.

A constant in many papers was the notion that investors, both retail and institutional, treated tech stocks as possessing more than their intrinsic value would suggest. Penman's paper on the quality of financial statements of tech companies (Undated) suggests that companies were being valued by metrics that were never before seen or used. The nature of the tech industry being "new" and "game-changing" led investors to use metrics such as page hits, bounce rates, and other hidden value factors to price companies instead of traditionally accepted measures with tangible effects on company strength such as earnings, dividends, and cash flow.

The phenomenon of attributing imaginary value to securities beyond their intrinsic range and into the realm of overvaluation is explored in other papers. In the paper written by Tuckett and Taffler, a methodical breakdown of investors' mental processes is presented (2005). Their conclusion is that investors went through a systematic phase of treating tech stocks as some sort of symbolic object that possessed unprecedented qualities and emanated a sense of hope for the future for their owners. One of the most important things touched upon was how egregiously overvalued certain securities were, and how they unrealistic were projections for future earnings growth if the efficient market hypothesis remained true. A paper written by Cooper, Dimitrov, and Rau goes in-depth describing some of the ludicrous examples of bubble mentality experienced during the time, such as huge price increases over meaningless news and information.[3]

Institutions were found to show signs of herd mentality, an approach to investing and trading supposedly reserved for only noise traders. Sharma, Vivek, Easterwood and Kumar (2006) prove the prevalence of institutional herding in the buildup to the tech bubble and find that it plays a significant role in the pricing of securities during the time. Valliere and Peterson look deeper into the subject of institutional investors from the Venture Capital field (2004), and they find evidence that also displays herd behavior. They find a strong correlation between unreasonably positive forecasts leading to excessive betting and leveraged order flow, causing valuations unbefitting intrinsic value and reasonable expectations.

Brunnermeier and Nagel's empirical evidence regarding hedge funds and their lack of moderation in fairly valuing securities (2004) is especially relevant; it shows that institutional investors did not act as contrarians in making sure that stocks were not overhyped, and they even had reason to want the irrational positive momentum of tech stocks to continue inflating the inevitable bubble. Nonetheless, the case for the institutional failure of market efficiency during the Tech Bubble becomes one worthy of attention and closer inspection.

Institutional Forethought?

As mentioned before, institutions are generally thought to be ahead of the curve. Empirical evidence suggests that this is true, but not in the way the efficient market hypothesis would have one believe. In other words, institutions were the first to act, but not in the interest of valuing stocks fairly. Hedge funds, a major representative of market participants with significant leverage in their capital, were some of the first buyers that tilted the valuation of tech securities to their intrinsic value. The funds then failed to act in a contrarian manner when retail and noise traders began to increase their holdings of them as well. As the bubble grew in size, institutions represented a larger portion of buying pressure than individuals.[4]

A common measure of market risk and investor irrationality is volatility. If people decide to exchange large numbers of shares in a short time, it is considered to be a sign of irrational behavior at the least. However, in this case, it can also be a sign to show that the efficient market hypothesis was violated. With such significant price movements as those created by the tech bubble, and with exorbitantly high valuations in the run-up and volatility created through a rapid sell-off afterwards, at some point there must have been pricing inefficiency. The efficient market hypothesis states that all securities are going to be properly valued with a reasonable sense of expectation for future cash flow and growth opportunities. However, the fact that an entire 65 percent of market capitalization of the NASDAQ Composite was shed in the deflation of the bubble strongly suggests the falsity of the statement. For the most part, it was unlikely that reasonable expectations for the future of publicly traded tech companies could have fallen to the point where the aforementioned amount of volatility could be justified.

In fact, individual retail investors actually were the ones to behave in a contrarian manner during the beginning of the bubble's popping. According to empirical evidence of order flow and holding disclosure, when the institutions began to rapidly sell off their shares and dump their positions in the securities they were supposedly fairly valuing, retail investors were the ones to behave in a contrarian manner. For a time after the selloff began, order activity from individuals showed buying activity while institutions were selling.[4] They had the mentality to "buy on the dip" as it is called, which is a strategy geared more towards long-term appreciation than riding short-term movements of volatility and irrationality. The ones actually mediating the spurt of selling activity, or at least making an attempt to, were individuals.

Institutions are not meant to be the parties initiating random buying pressure. They are supposed to be rational arbitrageurs, reacting to the inconsistencies and irrationality of individuals and noise traders to make sure that assets are valued fairly. In the case of the dotcom bubble, this was not the case. Now, the argument can be made that institutions were simply trading with the news, that they were the first ones to respond to new information that would significantly affect projections and valuations of tech securities. However, evidence also suggests that the largest institutions and banks traded technology stocks along with the movement of returns. [4] Basically, they employed a momentum-trading strategy, in which they tried to generate further returns by using past buying pressure and price movements as predictors for the future. This directly violates the underlying principles behind the efficient market theory, regardless of its various interpretations of strength.

Institutions using a momentum trading strategy is the polar opposite behavior from what an efficient market claims the smart money should do. The very nature of trading on momentum is entering a position because one believes the asset is "trending," through seeing that buying pressure has been growing quickly in a relatively short time. Basically, it is buying merely on the notion that others have *been* buying and will hopefully continue to

buy more. It is the essence of desiring market mispricing and trying to profit off of others' misguided emotions.

Lastly, the sheer scale at which institutions were exposed to the long side of tech stocks should be mentioned. There may still be evidence of institutional mispricing given that there were short-sale restrictions in place during the time. However, they were so far exposed that even short-sale restrictions cannot explain the overwhelming enthusiasm institutions showed towards the tech sector. The buying activity exhibited made short-sale constraints marginal in comparison.[2]

Irrational Exuberance to News

There is still the possibility, however, that institutions were simply ahead of the curve and processed the information before everyone else did. Both the dramatic price increase as the bubble formed, as well as the significant devaluation, could have been justified using fundamental data. However, there is significant evidence that refutes this claim. First, in Amazon's annual report for the fiscal year 2000, their business expanded and had an objectively better year in terms of their raw performance in numbers than the year before in various aspects of their business.[1] For example, sales grew to $2.76 billion in 2000 from $1.64 billion in 1999, and gross profit grew to $656 million in 2000 from $291 million in 1999, which is a 125% increase.

The stock performance of this company during that year is an anomaly. In the interest of not trying to completely discredit the efficient market hypothesis, the 80% decrease in share value of this company during the bubble's pop is arguably too significant to say that the price action of the company was justified throughout the entire time period of the tech bubble. Although it would be fair to say that the overall devaluation of the industry could have brought down Amazon's expectations, when a stock is losing 4/5ths of its value even after a truly good year, either the initial valuation was unjustified or the selloff was unjustified. If any of those statements are true, then the efficient market theory becomes nullified.

Also, examination of institutional order flow during days of significant news releases finds that there is still a positive correlation between price movements and institutional buying pressure on days *even* with little to no related fundamental news.[4] Therefore, the trading patterns of institutions was due not necessarily to their ability to track news and accurately price securities; they traded irrationally and bought tech stocks even though there was little to no reason to during some days. Again, the efficient market hypothesis stipulates that institutions are *always* trading accurately and fairly according to the fair value of financial assets. This was obviously not the case.

One's examination of news articles and market commentary that were widespread at the time show that news releases that had a palpable correlation with the underlying fundamental of stocks were actually sparse[8] Many times, raw facts and numbers were confused with out-of-reach predictions and wishful thinking. Companies released news that detailed their substantial increase in page-views, creative ways of monetization, and promotion of new services and processes. These claims, for the most part, were little more than fluff. Nonetheless, the public marketplace gladly ate it up and pushed valuations to stratospheric levels.

The media also reciprocated with positive news releases and commentary that served to further increase investor hype and make the technology sector ever more in demand.[9] The institutions themselves were promoting these securities with massive excitement, and in retrospect, severe negligence. An excerpt from Merrill Lynch's *Internet/E-Commerce Report*, published on March 9, 1999, claimed the following:

> "The overall Internet stock phenomenon may well be a 'bubble' but in at least one respect it is very different from other bubbles: there are great fundamental reasons to own these stocks ....The companies underneath are (1) growing amazingly quickly, and (2) threatening the status quo in multiple sectors of the economy ….With these types of investments, we would also argue that the 'real' risk is not losing some money – it is missing a much bigger upside."

This statement turned out to be hopelessly wrong. There is not only proof that institutions were at the very least unable to maintain a fair appraisal for the market price of tech companies, but also that they were flat-out wrong in determining their future prospects. They believed the future of tech companies during the bubble to be far greater than it really were and also publicly announced their supposed "fair valuation" that turned out to be completely wrong.

Unjustifiable Valuations

The underlying valuations themselves should have caused institutions to seriously evaluate their thinking and determine if the level at which they were pricing securities was justifiable. The average internet stock across the whole sector was valued on an implied P/E ratio of 605 in February 2000, and the required growth rate to maintain that valuation and deem it justifiable would be earnings of 63% yearly for 10 years. According to Tuckett and Taffler, this would have been more than "the top 2% of earnings performers from 1951-1958, across the whole internet sector."[9]

At some point, institutions need to be seriously questioned and held accountable for the rapid increase in valuation that occurred, as well as the irrationally exuberant expectations that were priced into the stocks. Most certainly, they should have shown some semblance of rationality to prevent the bubble from getting as large as it did. Nonetheless, evidence suggests the opposite took place. Correlational evidence suggests that institutions chose to forgo the involvement of certain necessary fundamental variables into their appraisal of securities during the lead-up of the Dotcom bubble.[6] It was only after the selloff began and the bubble popped that they began to incorporate all of the variables into the prices at which tech stocks were marked.

Tech companies with a web presence were not valued by the same standards as other sectors. New variables and ratios were being used, such as bounce rates, page views, and clicks per minute.[5] Commonly accepted factors used in security valuation such as dividend payout, cash flow, and earnings had no tangible relationship with the future financial health of the company. Even though these website-related variables were unproven and untested, they had so much weight on institutional perception of these companies that the companies became egregiously mispriced.

The behavior of institutions during the crisis was not completely unheard of, and it is in retrospect that one is systematic. The entire population of market participants, including retail and professional traders, treated Dotcom stocks as some sort of surreal phenomena that molded people's unconscious fantasies. According to a psychoanalytic study which examined investor behavior during the dot-com bubble, institutions went through a series of "compelling, exciting, and then terrifying and shameful emotions which dominated the thoughts of investors and of which they were largely unconscious."[9] Their behavior fit uncannily into a five-phased psychoanalytic theory, with those who had contrarian viewpoints or level-headed opinions being tossed aside and mocked. In fact, it became so absurd that in 1998 and 1999, even the addition of ".com" to the end of the name of a publicly traded company that had otherwise little to do with the internet would enjoy a

cumulative average return of 63% for the next five days after the announcement of the name change was released to the public.[3]

Institutions also exhibited considerable herding behavior regarding internet stocks during the Dotcom bubble. Analysis of holdings releases and of order flow shows that institutions entered into positions within tech stocks in a significant congregation and exited their positions in much the same way.[7] according to observation, tech stocks rose in some of the most significant amounts during the onset of maximum herd buying from institutions, and those very stocks dropped the most when the institutional herd buying behavior stopped.[7] Clearly institutions were directly at fault for a significant amount of the volatility and rapid price action movement that occurred throughout the bubble, an alarming finding when considering again that their role in an efficient market is supposed to be the ones that actively limit volatility and monitor unjustified emotional trading activity.

Moreover, some of the most poignant and significant evidence that disproves the efficient market hypothesis is a study during 1998-2001 that shows professionals at institutions personally agreeing with concepts that go against the very nature of their role in an efficient market. In an alarming majority, participants in a survey answered positively to statements such as "Market hype generates investor hype;" "Increasing valuations increase investor hype," and "Investor hype generates portfolio investing and flipping and increases sectoral premia."[10] The professionals, the contrarians, the ones that are supposed to be fairly valuing these securities, are admitting that the circumstances and behavior of the markets are at odds with efficient markets. Hype and increased speculation would be minimized if institutions are making sure that exuberance is not priced into companies, and that only reasonable expectations and grounded forecasts are.

Institutional investors admitted that if the markets truly were efficient, they did not successfully do their jobs, and this is not ever supposed to happen according to the principle behind an efficient market. The professional institutions are *always* supposed to be monitoring securities; there is no room for exceptions to be made. The theory states that stocks, or any financial asset, are always going to be accurately valued and marked no matter what. The overwhelming institutional recognition of investor hype in the financial markets makes the theory wrong. The markets were clearly inefficient.

Conclusion

There was a significant lack of disparity between institutional and retail trader behavior during the Dotcom bubble. The contrarian institutions meant to serve by remaining emotionless and fairly pricing securities, regardless of the extent of irrational activity by noise traders, did not do their job in the bubble's cycle. The efficient market theory became null and void during the bubble.

Evidence suggests that institutions bought before retail traders and sold before them as well. They contributed more to the volatility of the markets than individuals and rode the Dotcom bubble just like them. There was no evidence of significant fundamental data that could affect company valuations in the lead-up to the Dotcom bubble or its collapse. The sheer volatility present during the bubble's cycle is evidence of emotion rather than efficiency taking hold of the market.

Also, the implied growth rates in earnings according to company valuations were far too optimistic to be feasible. Institutions also behaved in a systematically emotional way and admitted themselves that inefficient market tendencies and behaviors pervaded the market. Lastly, there was a significant deviation in value from what could be considered feasible given the economic data present at the time. Therefore, the only fair conclusion to

make is that the efficient market theory was not applicable to the financial markets during the cycle of the Dotcom bubble.

In no way does this paper condone the derision of professional traders and portfolio managers that work at financial institutions. They are accomplished in their own right and have spent quite a bit of their time perfecting their craft. Again, to say that their college educations and years of experience maneuvering the markets are meaningless and do not add to their ability to outperform the overall population of market participants would be forgoing common sense and echoing unsubstantiated emotional opinion that is simply not legitimate.

However, this paper is expressing the possibly surprising lack of discrepancy between professionals and amateur noise traders during the Dotcom bubble. It is important for *all* market participants to remain vigilant that their investments and positions are not tainted by erratic market emotion. Of course, identifying emotion is extremely hard to do in practice, but leaving the job to institutions and professionals under the guise of theoretical application can cause significant turmoil in the markets as well as cataclysmic losses that all investors endured in the wake of the collapse. Institutions are comprised of people, too. There is no animosity being expressed here, just a restatement of a cold, hard fact that some inevitably forget in periods of extreme euphoria and exuberance.

It is important to remember that this money was *real* money, and that these losses were *real* losses. Individual investors lost huge sums of savings in the aftermath of technological economy's collapse. Professionals and institutions cannot be regarded as the protectors of market neutrality and fairness. The possibility of losing or gaining exorbitant sums of money from market inefficiency and bubble creation supposedly under their watch was very real. And if history has taught us anything, it is that bubbles will inevitably occur again at some point or another.

References

[1] Amazon. *2000 Annual Report*, 2001. Web. 14 Aug. 2016.
[2] Brunnermeier, Markus K, and Stefan Nagel. "Hedge Funds and the Technology Bubble". The Journal of Finance 59 (2004): , 59, 2013-2040. Print.
[3] Cooper, M. J., Dimitrov, O. and Rau, P. R. (2001). A rose.com by any other name. Journal of Finance, 56(6), 2371-2388.
[4] Griffin, John M., Jeffrey H. Harris, Tao Shu, and Selim Topaloglu. "Who Drove and Burst the Tech Bubble?" *The Journal of Finance* 66.4 (2011): 1251-290. *JSTOR*. Web. 13 Aug. 2016.
[5] Penman, Stephen H., The Quality of Financial Statements: Perspectives from the Recent Stock Market Bubble (Undated). Available at SSRN: http://ssrn.com/abstract=319262 or http://dx.doi.org/10.2139/ssrn.319262 Web. 14 Aug. 2016.
[6] Peter Robert Wheale & Laura Heredia Amin (2003) Bursting the dot.com "Bubble': A Case Study in Investor Behaviour, Technology Analysis & Strategic Management, 15:1, 117-136,
DOI: 10.1080/0953732032000046097
[7] Sharma, Vivek, John Easterwood, and Raman Kumar. "Institutional herding and the internet bubble." Unpublished working paper, University of Michigan-Dearborn and Virginia Tech (2006).
[8] Thomas G. Goodnight & Sandy Green (2010) Rhetoric, Risk, and Markets: The Dot-Com Bubble, Quarterly Journal of Speech, 96:2, 115-140,
DOI: 10.1080/00335631003796669
[9] Tuckett, David, and Richard J. Taffler. "A Psychoanalytic Interpretation of Dot.com Stock Valuations." *SSRN Electronic Journal* (2005): n. pag. *SSRN*. Web. 13 Aug. 2016.
[10] Valliere, Dave, and Rein Peterson. "Inflating the Bubble: Examining Dot-com Investor Behaviour." *Venture Capital* 6.1 (2004): 1-22. *Taylor & Francis*. Web. 28 Aug. 2016.

Characterization of Chitosan/PVA Scaffolds with Chitosans of Different Average Molecular Weights for Tissue Engineering

Zijun Zhang

Author background: Zijun Zhang grew up in China and currently attends The Experimental High School Attached to Beijing Normal University, located in Beijing, China. Her Pioneer seminar topic was in the field of chemistry and titled "Glycoscience: From Materials to Medicine."

1. Introduction

1.1 Tissue Engineering

Tissue engineering, the technique used to repair damaged or lost tissues and organs by replacing them with laboratory produced multi-tissue organs, tissue interfaces and structural tissue and muscles, is becoming an increasingly popular field with a promising future. For in vitro tissue engineering, it is important that proper three-dimensional scaffolds can be created to guide cellular growth and proliferation, and to support new tissue formation. *(1)* Many of the first scaffolds were made by nondegradable materials so operations were needed to remove the scaffolds after the tissues and organs finished growing. However, this process was tedious and added risks. This led to the development of degradable scaffolds. In original studies, many naturally derived sugars like cellulose, agarose, alginate and chitosan were selected as possible materials for degradable scaffolds because of their biocompatibility, low cost and ease of processing. *(2)* Later studies demonstrated that chitosan was more suitable due to some of its distinctive properties.

1.2 Chitin, Chitosan and Chitosan/PVA

Chitosan is a derivative of chitin, a polysaccharide composed of *N*-acetylglucosamine linked by ß-1,4 glucosidic bonds. Chitin is the main component of the exoskeletons of crustaceans (like crabs and shrimps) and insects, as well as the cell walls of many fungi. Therefore, it is abundant in nature and easy to possess. Chitosan is obtained by deacetylation of chitin and is normally defined as a copolymer of *N*-acetylglucosamine and glucosamine with degree of deacetylation (D.D.) higher than 60%. Unlike chitin which has poor solubility, chitosan is soluble in dilute acid and mildly acidic solution, which enables its various applications.

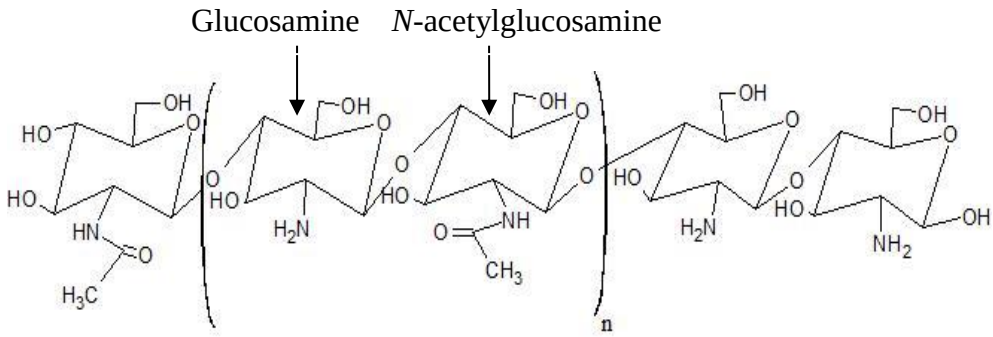


Figure 1. *Structure of a chitosan molecule.*

Chitosan's excellent biocompatibility, bioactivity, biodegradability, nontoxicity and antimicrobial and antioxidant activity enable its versatile applications in food, agriculture and biomedicine. *(3)* However, its weak mechanical properties are also problematic. Due to its high brittleness and low tensile strength, it lacks the stability to oppose the stresses induced during cellular growth. To solve this problem, synthetic polymers have been used in combination with chitosan to form hybrid materials. Some of the materials that have already been produced include chitosan with mixtures of poly(ethylene oxide), poly(caprolactone), poly(vinyl alcohol), poly(acrylamide), PET, poly(amidoamine), poly(*N*-isopropylacrylamide), poly(glycolic acid), poly(lactic acid), poly(ethylene glycol), poly(acrylonitrile), poly(ether block amide), poly(propylene carbonate) and poly(y-glutamic acid). *(4)*

While many of the aforementioned hybrid materials can be used to produce thin films for applications in food industry and agriculture, only poly(vinyl alcohol) (PVA) has shown promise in the biomedical field (especially in wound healing, wound dressing and tissue engineering). One reason for this is because of its highly hydrophilic properties, excellent mechanical properties, nontoxicity, water solubility, biodegradability, biocompatibility and its low cost. *(5)* Chitosans and PVAs are made into hydrogels for biomedical uses and the main interactions in the structure are hydrogen bonds between chitosan's amino groups and PVA's hydroxyl groups. Since hydrogels can swell in liquid solution and retain a significant fraction of water in their porous structure without dissolving, they have physical properties similar to human tissues, which means they have excellent tissue compatibility. *(6)* Normally, chitosan/PVA hydrogels used as scaffolds are produced by autoclaving method.

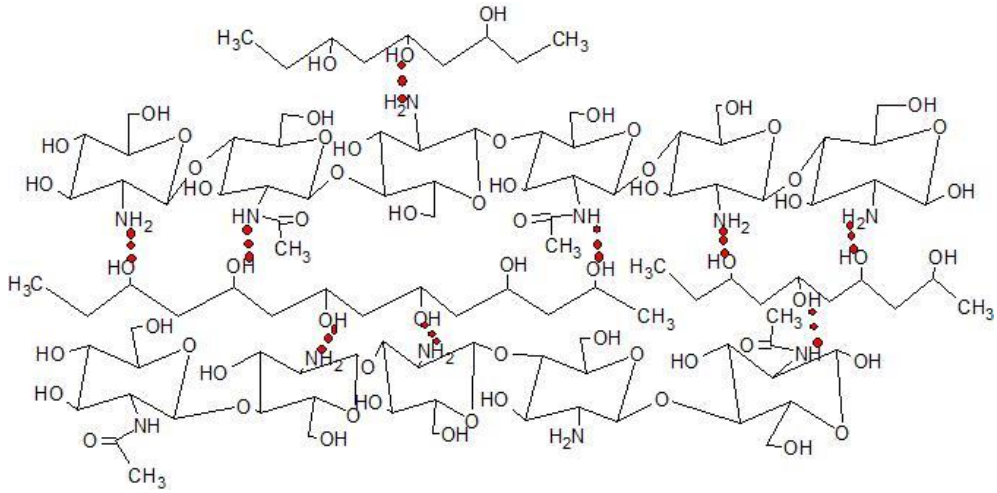


Figure (2) *The interaction between chitosan and PVA molecules through hydrogen bonds*

1.3 Previous Studies

In recent years, several studies based on chitosan/PVA three-dimensional scaffolds have been conducted and each are stimulating further development to improve the function of the scaffolds. For example, experiments using chitosan/PVA membranes of different chitosan to PVA ratios have demonstrated that the membranes have strong hydrogen bonding interactions between the two kinds of molecules, thermal stability, improved mechanical properties, and excellent wettability. In vitro biocompatibility tests have also shown that cultured cells can stick to the surface of the chitosan/PVA mixture and grow rapidly without microbial infection. These results demonstrate that while pure PVA can neither promote cell growth nor inhibit microbial infection, the addition of PVA into chitosan can make up for the disadvantages of chitosan, and chitosan can still stimulate cellular growth as effectively as when no PVA is added into the scaffold.

Another study blended chitosan/PVA with methylcellulose to produce a new scaffold. Researchers found that the addition of methylcellulose increased the porosity and improved the morphology of the scaffold, which means that there would be more evenly distributed pores in the scaffold to deliver nutrition to the cultured cells. *(7)* Apart from that, calcium carbonate was used in another study to examine whether it could enhance the mechanical properties of the chitosan/PVA scaffold. Experimental results showed that the highest ultimate stress value of the scaffold was largely increased and the changes in biocompatibility could be ignored. *(8)*

Studies above show that several extra materials have been studied to improve different aspects of the scaffold, which also indicates that weaknesses still remain for the pure chitosan/PVA scaffolds. Therefore, improvements still need to be made to address the weaknesses of these scaffolds. The research in this proposal attempts to address this issue by studying chitosans of certain molecular weights in order to find out if it is possible to improve certain properties of the chitosan/PVA scaffolds.

1.4 Research Purpose and Hypothesis

Although the addition of both methylcellulose and calcium carbonate improves certain properties of the chitosan/PVA scaffold, it simultaneously reduces the concentration

of chitosan in the scaffold. Chitosan is the only effective material in the scaffold that enhances cell growth by providing tissue-like surfaces with nutrient transporting pores and by preventing microbial infection. Therefore, reduced concentrations of effective chitosans may hinder the scaffolds from fully functioning. It is necessary to find a way that both maintains chitosan's effects and keeps the properties of the scaffold suitable. In this research, chitosans with different average molecular weights will be used to produce chitosan/PVA scaffolds to achieve the goal mentioned above.

Previous studies investigated the relation between chitosan's molecular weight and antibacterial effect. A study on the influence of chitosan's degrees of acetylation and molecular weights on its antibacterial and antifungal activities have shown that chitosans with a higher degree of deacetylation (D.D.) and a smaller molecular weight increase those activities. *(9)* However, in another study on chitosan-impregnated bacterial cellulose film, chitosan with higher molecular weights proved to have better antimicrobial activity, which contradicted the results in the study mentioned earlier. *(10)* In this study, chitosan with lower molecular weight appeared to promote human cell growth better. *(10)*

As is shown, studies on pure chitosans and mixtures containing chitosans are showing conflicting results regarding molecular weight. Nevertheless, neither of them was conducted on chitosan/PVA and neither took the changes of other properties like tensile strength and porosity into consideration, so what will be caused by molecular weight difference of chitosan in chitosan/PVA scaffold is still unknown. However, since the former study showed the benefit of chitosans with high D.D. and other studies on chitosans also prefer high D.D. ones, it is reasonable that chitosans in this study also have high D.D.

In this study, chitosans of different average molecular weights (MW) (from 42.5 to 135 kDa) with the same degree of deacetylation (92%) will be generated through an enzymatic approach and then undergo a purification process. The resulting chitosan fragments will then be blended with PVA to form chitosan/PVA three-dimensional scaffolds with chitosan:PVA ratios of 1:1. These scaffolds will be examined for their compatibility, thermal stability, mechanical properties, porosity, morphology, wettability, biocompatibility and biodegradability.

1.5 Expected Result

This study may have several possible results. Scaffolds with chitosans in certain average molecular weight ranges may display much stronger antibacterial effects so that using these chitosans with relatively lower concentrations in the scaffold will not have a conceivable influence on the growth of the tissue. Another possibility is that some scaffolds may prove to have greater interactions with PVA, or have larger porosity, so that fewer additional materials will be needed and the concentration of chitosans will be increased. Though there are different possible results, in both cases the effectiveness of chitosans can be emphasized. Hopefully, the study results can help the scaffold be applied to produce more kinds of tissues and organs.

2. Experiments

2.1 Preparation

2.1.1 Preparation of chitosans of different average molecular weights

The original chitosans used to produce chitosans of different molecular weights (MW) will be commercially available. They will have the same MW with the same degree

of deacetylation (92%). These chitosans will then undergo enzymatic cleavage. The preparation will follow Xie Y.'s enzymatic hydrolysis method, using cellulose from *A niger*. to hydrolyze the original chitosans into chitosans of different MWs ranging from 43.5 kDa to 135 kDa. *(11)* Those chitosans with different MWs will finally be purified by first dissolving them in acetic acid solution, then filtering under diminished pressure and finally forming precipitates in concentrated NaOH solution. *(11)*

2.1.2 Scaffold formation

Kanimozhi K.'s lyophilization method will be used to form scaffolds. *(7)* One mole of each kind of chitosan will be dissolved in 100 ml of 0.1 M acetic acid solution. These solutions will each be mixed with 100 ml water that contains 1 gram of PVA to form the original scaffolds. After freezing them at -20 degree Celsius, the final dried porous scaffolds will be obtained.

2.2 Characterization of scaffolds

2.2.1 Fourier Transform Infrared (FTIR) spectroscopy

FTIR spectroscopy will be used to characterize the hydrogen bonds between chitosan and PVA molecules in scaffolds formed by using chitosans of different MWs. FTIR spectroscopy is a method of characterizing special chemical functional groups in materials by using infrared radiation. The infrared radiation induces bond vibrations in materials and since the patterns of vibration are different for different chemical groups, compounds and groups of compounds, their spectra will also be unique. In the final spectrum, each stretch can be traced to a specific functional group. However, the pattern and shape of the stretch will change according to whether or not there are interactions between specific functional groups.

In this proposal, Mansur H.S.'s method will be followed in this test to obtain the spectra of PVA molecules' hydroxyl groups in each scaffold. *(12)* Comparing with the spectra of standard PVA molecule, PVA molecules that have hydrogen bonds with chitosans have a wider stretch. Therefore, by comparing stretch width of the spectrum of all those hydroxyl groups in PVA molecules, the test will figure out the amount of hydrogen bonds in each scaffold, which indicates the strength of chitosan/PVA interaction.

2.2.2 Testing of mechanical properties

Mechanical properties of different scaffolds will be evaluated by a tensile test. The tensile test measures the maximum pressure that can be induced on an object before its failure. Tensile strength measurements of each kind of scaffold will be obtained using Zhuang P.Y.'s method.*(5)* Each kind of scaffold will be stretched to a maximum length by a controlled tension so that data of the maximum elongation will be gained. At least five samples of each kind will be tested. Comparing the maximum elongation of each kind of scaffold, the scaffolds with enough tensile strength to support the attaching cells will be selected.

2.2.3 Testing of swelling degree

Each scaffold will be measured for its degree of swelling. Swelling degree measures the amount of water each scaffold can absorb. Scaffolds with higher swelling degree can hold more nutrients to support cell growth. Kanimozhi K.'s method will be used to test the swelling degrees. *(7)* One sample of each scaffold will be soaked in a phosphate buffer solution for 30 days so that each sample will take up the maximum amount of solution. By measuring the original weight and the weight after soaking of each sample, the percentage of water absorption will be calculated for each sample. Those samples with the percentages of water absorption around 5% will be considered to have the suitable swelling degree for the scaffolds to hold nutrients.

2.2.4 Testing of surface morphology using scanning electron microscopy (SEM)

SEM microscopy provides the images of samples by having electrons interact with atoms of the samples to create various signals indicating sample's surface morphology. In this proposal, SEM microscopy will be used to test the distribution of pores on surfaces of each scaffold. It is expected that scaffolds with more evenly distributed pores will be selected from the others with this method since even distribution of pores enables each cell on the scaffold to obtain the same amount of nutrition so that the cells will all grow well at the same speed.

2.2.5 Testing of wettability

The wettabilities of each scaffold type will be measured by running a contact angle test. The contact angle of a scaffold and water is the angle that water meets the surface of the scaffold. A lower contact angle implies the hydrophilic property of certain scaffold, which means that the scaffold has better wettability, so that liquid can better interact with the surface and later be absorbed.

In this test, Zhuang P.Y.'s method will be followed to test the contact angles of different chitosan/PVA scaffolds. A drop of distilled water will be dropped on five different parts of the surface of each kind of scaffold. The contact angles between the surfaces and the water drops will then be measured by a contact angle goniometer. Samples with the highest contact angles will be considered to be most suitable. *(5)*

2.2.6 In vitro tests

2.2.6.1 Testing of degradation

A proper speed of degradation is important for the scaffold so that it can finish degradation just as cell growth completes. The rates of degradation can be measured by obtaining the percentages of weight loss of each sample following Kanimozhi K.'s method. *(5)* Each scaffold will be immersed in Hank's solution (simulated body fluid composed of Na^+, K^+, Ca^+, Cl^-, HCO_3^-, $H_2PO_4^-$, and glucose) for a given time. The weights of each sample before and after immersion will be used to calculate the percentage weight loss. The rates of degradation for all samples together with the rates of cell proliferation on all scaffolds (see section 2.2.6.3) will help select the proper scaffolds whose rates of degradation correspond to the cell growth.

2.2.6.2 Bacterial assay

The antimicrobial effect of each kind of scaffold will be tested by running a bacterial assay. Scaffolds with better antimicrobial effects will inhibit microbial growth so that growing tissues will not be infected. Same size samples of each scaffold will first be sterilized. Each scaffold sample will then be inoculated with *E coli.* and *S aureus*. separately for a given time. The counts of each kind of bacteria on each sample will be obtained. The samples with the smallest bacterial counts will be selected.

2.2.6.3 MTT assay

The cell proliferation rate for each kind of scaffold will be tested using MTT assay. MTT is a water soluble yellow dye which can be reduced to purple formazan by mitochondrial reductase. The amount of mitochondrial reductase in a sample is directly proportional to the number of living cells in a given sample and a larger amount of the reductase leads to a darker purple color. Therefore, the darkness of the purple color of each sample indicates the number of living cells present. The darkness of each sample can be quantified by measuring with UV-visible spectroscopy.

In this proposal, bovine serum albumin cells will be used for MTT assay using Kanimozhi K.'s method. *(7)* Same number of cells will be incubated on samples of each scaffold in 96 well plates for 24 hours. MTT will then be added to the plates. Cell proliferation rates will be obtained by measuring the light absorptions of the samples in 1, 4 and 7 days. Samples with highest absorptions, that is, highest cell proliferation rates, will be considered the most suitable.

3. Conclusion

In order to find a method to build suitable chitosan/PVA three dimensional scaffolds for tissue engineering without using additives that reduce the effect of chitosans, chitosans with different molecular weights will be tested to form scaffolds in this proposal. Chitosans with different molecular weights will be produced by enzymatic cleavage. They will then be purified and blended with PVA to form three dimensional scaffolds for tissue engineering.

The properties of chitosan/PVA scaffolds with chitosans of different molecular weights will be characterized using FTIR spectroscopy, tensile test, swelling degree test, SEM, contact angle test, degradation rate test, bacterial assay and MTT assay. FTIR spectroscopy will be used to examine the amounts of hydrogen bonds in different scaffold, which indicate the strengths of chitosan/PVA interaction. Tensile testing will reveal the maximum elongation of each different scaffolds, which will be used to determine which scaffolds have the best mechanical properties to support the attaching cells. Swelling degree tests will be used to determine the percentages of water absorption for each scaffold and help select scaffolds that hold the most nutrients for cell growth. SEM will provide images of the surface of each scaffold, showing the distribution of pores on each scaffold. Scaffolds with more evenly distributed pores, which allow cells to well absorb nutrients and grow at the same speed, will be selected by this method. Contact angle tests will measure the wettabilities of different scaffolds so that scaffolds that easily interact with liquids will be selected because they allow nutrients in the solution to be better absorbed. By running an in vitro degradation test using Hank's solution, the percentages of weight loss for different scaffolds will be obtained. These data together with the data of cell proliferation rates will be used to determine which scaffolds have the most proper rates of degradation that correspond to the rate of cell growth. *E coli.* and *S aureus.* bacterial assays will be used to

determine the scaffolds with the best antimicrobial effects. Finally, the MTT assays will be used to provide information on which scaffolds have the best cell proliferation rates.

The chitosan/PVA scaffolds with chitosans of certain molecular ranges that behave well in the majority of experiments outlined above will be considered the most suitable scaffolds for tissue engineering. Hopefully, the suitable scaffolds will be produced with chitosans of certain molecular weight ranges and will replace scaffolds that contain significant amounts of additives in tissue engineering.

4. References

(1) Murugan R.; Ramakrishna S.; *Tissue Engineering* **2006**, *12*, 435-447
(2) Ko H.F.; Charles Sfeir C.; Kumta P.N. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A* **2010**, *368*, 1981-1997
(3) Park B.K.; Kim M.M. *International Journal of Molecular Sciences* **2010**, *11*, 5152-5164
(4) Rafique A.; Zia K.M.; Zuber M.; Tabasum S.; Rehman S. *International Journal of Biological Macromolecules* **2016**, *87*, 141-154
(5) Zhuang P.Y.; Li Y.L.; Fan L.; Lin J.; Hu Q.L. *International Journal of Biological Macromolecules* **2012**, *50*, 658-663
(6) Yang J.M.; Su W.Y.; Leu T.L.; Yang M.C. *Journal of Membrane Science* **2004**, *236*, 39-51 (7) Kanimozhi K.; Basha K.S.; Kumari V.S. *Materials Science and Engineering: C* **2016**, *61*, 484-491
(8) Sambudi N.S.; Park S.B.; Cho K. *Journal of Materials Science* **2016**, *51*, 7742-775
(9) Younes I.; Sellimi S.; Rinaudo M.; Jellouli K. Nasri M. *International Journal of Good Microbiology* **2014**, *185*, 57-63
(10) Kingkaew J.; Kirdponpattara S.; Sanchavanakit N.; Pavasant P.; Phisalaphong M. *Biotechnology and Bioprocess Engineering* **2014**, *19*, 534-544
(11) Xie Y.; Hu J.; Wei Y.; Hong X. *Polymer Degradation and Stability* **2009**, *94*, 1895-1899
(12) Mansur H.S.; Sadahira C.M.; Souza A.N.; Mansur A.A.P. *Materials Science and Engineering: C* **2008**, *28*, 539-548

Athena's Spoiled Olives-
How Institutional Flaws of the European Union and Greek Politics Shaped a Failing Economy

Sean Hu

Author background: Sean Hu grew up in Taiwan and currently attends Pacific American School, located in Hsinchu City, Taiwan. His Pioneer seminar topic was in the field of international relations and titled "Political Institutions of the World."

Note

To provide better visualization for the contents of this research inquiry, the information and data examined have been presented in two different mediums. The first is the ESRI Story Map, which integrates the content of the research with the corresponding visuals that can come in the form of videos, images, or even interactive maps. Using this format, each section of the essay, with the addition of a succinct introduction of the establishment of the European Union, is briefed in the story map. Infograms, on the other hand, are more proficient for presenting graphs and statistics, which are followed by a short explanation. Therefore, each figure, with its supplementary data, was accumulated and displayed in the infogram. The accesses to both are provided below in the footnote[53].

Introduction

Within three decades of its formal establishment by the Maastricht Treaty, the European Union (EU) has seen numerous successes in consolidating integration. Nevertheless, it has also faced several challenges that have impeded its progress to forge a politically and economically united entity. Perhaps the most fitting illustration of such a hurdle, which has threatened the Union's stability, pertains to the crisis in Greece. Greece offers a unique glimpse at the structural flaws of the economic institutions that comprise the EU. These flaws culmination in the Greek sovereign debt crisis and barred the EU from effectively addressing the issue. Coupled with its disadvantageous economic model, which reflected the political infighting of the country in the 1990s and 2000s, Greece contributed to the surge of backlash against the EU's integration efforts. This pressing situation challenged both the survival of the Greek economy in the EU and the monetary union as well as the stability of those institutions. In preparation for the analysis, we must first consider the following question: How did the structure of the institutions of the European Union lead to economic and political difficulties in Greece as well as the development of a backlash against integration from its members?

Understanding the international relations with regards to Greece in the context of the European Union demands a multifaceted approach. This approach requires analyses of the political, social, and economic aspects. Therefore, this research assignment aims initially to

[53] **Story Map:***http://arcg.is/29VMl80*
Infogram: *https://infogr.am/45aa7efa-f5d8-4f1c-a786-2486b8a72c9f*

frame the context of Greece as a part of the European Community/Union. The analysis subsequently explains that the crisis was a product of the Greek government's neglect of responsibilities and the institutions of the Union's flaws. This is demonstrated through Greece's participation in the institutions of the EU—the Single European Market (SEM), followed by the European Economic and Monetary Union (EMU) over a decade later—as well as its political interactions, economic approaches, and the social responses which followed. This paper further demonstrates, through demographic breakdowns and a comparative analysis, that the Greece situation should not elicit fear of an economic spillover dreaded by the backlash which threatened to tear the EU apart. Instead, this crisis is a call for Greece to reform its economic and political priorities to be compatible with the actions the EU should take to restructure its institutions in order to safeguard the Union from imprudent behaviour of governments.

Throughout the research process, it was necessary to consult a variety of sources in order to find the evidence to support the project's claims and reasoning. The economic aspects of the essay employed primary sources such as governmental or institutional statistics and reports to show how trends developed over time. They include those published by the National Statistical Service of Greece, OECD, and the World Bank. Therefore, many of the inquiries are dependent on document analysis, a form of qualitative research using objective accounts of facts to produce the reasoning behind the conclusion. Some of the economic concepts were derived from the writings or statements of economists and investors such as Ben Bernanke and Asher Edelman. They provide a better context and more validity for the claims made regarding both the Greek and EMU economic structures. Regarding the political aspects, the research mainly focused on connecting the information provided by academic journals and writings with reports issued by both the European Commission and other governments. Of the most value, however, was an online interview with Dr. Anna Visvizi, a leading scholar on Greece and the European Union, made possible by Professor Colette Mazucelli. Given Dr. Visvizi's expertise in the Eurozone crisis, her unique insight allowed the analysis to extend beyond the information offered by the mainstream literature. Overall, secondary sources were the foundation for much of the research, while primary sources verified or strengthened those materials. For better visualization, the figures provided have been gathered and integrated into a collection of infograms, presented alongside an interactive story map which also offers supplementary information[54].

I. The Greek Context

Adjunct to the Second Enlargement of the larger EU, Greece first joined the European circle in 1981, alongside other Mediterranean countries, including Spain and Portugal which followed in 1986. The previous decade saw the end of several dictatorships which were followed by movements towards left-wing governments that supported democracies. From the standpoint of the European Communities[55], this tactical expansion was meant to secure Southern Europe from being absorbed into the Communist sphere of influence which could have allowed the Eastern-bloc to envelop the continent. Greece was unhesitant in its attempt to join the European Communities, as it tried to distinguish itself as a "European country"

[54]The link to the story map which summarizes the content of each section with an assortment of visuals to provide a more interactive presentation of information: *http://arcg.is/29VMl80*

[55] The European Communities referred to the European Coal and Steel Community (ECSC), EURATOM, and European Economic Community (EEC), prior to the establishment of the European Community which merged the three organizations

(Lesser, Larrabee, Zanini, and Vlachos-Dengler, 2001, 20-21). For years, Greece had suffered from right-wing military rule and a political system with weak institutional structures that allowed for the development of corruption and other abuses of power (Visvizi,2013, p. 26). These afflictions left the country distant from the democratic administrations which defined Europe. Therefore, the movement towards integrating with Europe built onto a desire to institute a more democratic system that was shared by both the centre-right- leaning New Democracy (ND) party and its counterpart on the other side of the political spectrum, the Panhellenic Socialist Movement (PASOK) government. However, discussions to join the European Economic Community (EEC) began as early as the 1960s with the Association Agreements which outlined cooperation between the European Communities and countries that were not members of the institutions. Nevertheless, the desire for 'democratization' of Greece was certainly strengthened as a result of the military junta and flawed regimes which followed. In a sense, reinstating Greece's position in Europe was to redefine the country's geopolitical prospects, and less for economic purposes, to be discussed in the following paragraph (Visvizi, 2013, p. 27) [see *Acknowledgments* section at the end of the paper].

Their desire to join the European Communities was demonstrated by applying just one year after the overthrow of military rule. Looking at Figures 2, 3, and 4, it was obvious that economic growth came with the integration into the European scene. The graphs of Greece's GDP and GDP per capita from 1960 to 2014 and Purchasing Power Parity (PPP) per capita from 1990 to 2014, respectively, have an undeniably similar trend—slow, bordering on flat growth and a turning point, which initiated significant growth before peaking around the same time. These junctures occurred when the economy saw a massive boost and reflected the years in which major events took place. They transformed the European Community into the Union that would dominate the European narrative for decades to come. Interestingly, the spikes did not come immediately upon Greece's entry into the European Community in 1981. Instead, GDP growth followed, to a smaller extent, the implementation of the SEM in 1987 and, more importantly, the introduction of the Euro in 2001 when Greece joined the third stage of the EMU. The same can be observed with GDP per capita and the PPP per capita following the official establishment of the EU under the Maastricht Treaty in 1992. This delayed jump translates to two conclusions: 1) Greece's original goal in joining the European Community was primarily political, as it reaped no significant economic milestones in the initial years; rather, it secured a seat in the European spotlight and ensured a movement towards democratization; and 2) prior to joining the Community, Greece saw meagre growth, requiring the EU to kick-start its economic expansion.

Nevertheless, there has been ubiquitous criticism of the Euro being harmful to the Greek economy because Greece, with a relatively weaker and more vulnerable economy, could not respond with changes in its monetary policies during an economic crisis. Furthermore, its fiscal behaviour would have to be dependent on the decisions of the Eurozone for the economy to be sustainable because monetary and fiscal policies must be compatible and Greece only has autonomy over one. Essentially, even without a fiscal union, they had already lost significant autonomy, given that their fiscal activities were subject to the changes of the European Central Bank (ECB) centred in Frankfurt, Germany. Yet, in spite of the numerous criticisms, the euro and Greece's participation in the EMU had striking beneficial effects on the Greek economy that led to unprecedented growth. As Figure 5 shows, in the decade and a half following the signing of the Maastricht Treaty, the annual GDP growth rates have stayed healthily above zero and with no hints of an impending economic crash. For many of the years, the growth even peaked and reached

staggering heights—up and beyond five percent—far over the percentages that are marked by the capacities of developed countries. This evidence is further corroborated by Figure 6, which is a comparative graph. Regarding the quarterly economic growth, Greece far outmatched the Eurozone as a whole, averaging 3.9 percent annual growth for a decade beginning from 1996, while the remainder of the Eurozone countries lagged behind at an average of 2.2 percent. In addition, the nationwide prosperity was bolstered by a significant increase in living standards as GDP per capita rose by 47 percent within the same timeframe.

However, before these successful changes could occur, Greece had to be further accepted into the EMU, in spite of its then-current membership. Beyond the establishment of the European Union, the Maastricht Treaty of 1992 also served a preparatory purpose as the continent shaped itself for the ensuing single currency. This groundwork came in the form of the 'convergence criteria,' a set of benchmarks for a country to reach before being able to become a part of the Eurozone established in the Treaty of the European Community Article 121(a), in order to ensure the institution's success (EUR-Lex). There was a strong desire to accomplish these goals in order to shift to the single European currency, as there should have been. Indeed, the strength of the euro was backed by a monetary union which consisted of almost half of the world's top ten leading economies, and therefore saw favourable economic growths that are demonstrated by the figures. In order to be qualified as a member of the Eurozone, Greece required massive economic improvements. Under the drachma, Greece was tormented by both high inflation and extreme currency fluctuations. Figure 7 clearly shows that from the end of the military junta in the 1970s, the inflation rates remained high and unstable. The rates even reached around 35% and dropped immediately to just above 10% in the following years, before climbing back. While the ideal inflation rate may differ among countries, the optimal level seems to be 2%. For almost three decades, from the early-1970s to just before the 2000s, Greece marked numbers far beyond 2%. This made the banking sector practically nonexistent because: 1) it had no incentive to loan the money out given that the returned money would be worthless unless the interest rates were unreasonably high; and 2) the banks would have had no customers because money sitting in the banks would only lose value, which resulted in people spending it immediately. The fluctuations contributed to a lack of stability of consumer and business confidence. These factors explain the high nominal interest rate the year the 'Maastricht criteria' was drafted, at a hefty 24.13 percent (European Commission). Both, however, are listed in the agreement as criteria one and four, respectively. To address the situation, the government engaged in heavy deficit spending policies to jumpstart the economy. The inflation rate, evidently, dropped substantially, and by 2001, Greece had met the prerequisite. The same was seen with nominal interest rates which, by 2001, lowered to 5.1 percent (OECD, 2016). The struggle continued as the second criteria was regarding controlling government finance, which meant low annual government deficit—not exceeding 3 percent—and ratio of gross government debt not exceeding 60 percent at the end of the preceding financial year (EUR-Lex). Thus, despite the balancing of deficit spending and lowering inflation/interest rates, this method was contradictory and the truth of the country's finances was falsified, only to be revealed later in 2004, when Greece had already merged with the EU. In conclusion, while there is indisputable evidence showing the economic growth Greece experienced was rewarding, the initial demands to be integrated into the single currency were heavily burdensome and marked the beginning of the Greek economy's obscured decline into the sovereign debt crisis.

II. Greek Participation in the Institutions of the EU

In terms of identifying the chief impacts of the Greek sovereign debt crisis, the two primary bodies within the EU to be taken into consideration are the SEM and the EMU. Similarly, the analysis of the issue is twofold: it first looks at the build-up towards the crisis, and continues by examining the response that followed. The nature of the Greek economic crisis is its sovereign debt; thus, the focus would be on the Greek deficit and its accumulation. Again, this category can be broken down into two further branches—Greece's trade deficit and its excessive borrowing. In regard to the trade deficit, Greece had several disadvantages that contributed to its decline, in spite of its involvement in the same institution that granted it substantial economic growth. With the SEM in place, the EU sanctioned the free movement of goods, services, and people. Coupled with the single currency, economic interaction within the EU and between Eurozone members came easily. This provided the means for the growth that Greece experienced throughout the 1990s and the 2000s. However, it also provided a peculiar burden for the country by driving up the cost of labour. Figure 8 depicts this gradual trend. While the graph is marked with drastic, small changes, a line of best fit clearly demonstrates a continuous increase from the first year of the data, 1995, to as late as 2012, when the pattern finally turned. As other countries in Europe opened up, better job opportunities in the economic core of the EU, such as Germany and France, provided alternatives for these workers. Furthermore, the higher wages in the more developed countries forced Greece to gradually match those rates to be competitive within the SEM and attract labour. This extends from the concept of the uneven spread of the benefits of integration. Not only is Greece a geographically peripheral country, it is also economically less influential. Its smaller economy and lack of sufficient comparative economic advantages made it difficult to provide the necessary wages needed to attract workers (Magoulios and Athianos, 2013, p. 208).

Structurally, the SEM is disadvantageous to economically peripheral countries because barriers that could be employed to ensure a favourable balance of trade are forcibly removed. Figure 9 demonstrates that since it joined the Eurozone, Greece has sustained a negative balance of trade overall and it steadily dropped until the crash forced more beneficial terms from other EU countries, in order to ensure the survival of Greece's economy. Ironically, despite the efforts Greece exerted to join the EU and become a part of the Eurozone, Figure 10 shows that Greece did not choose to exploit the benefits of the SEM, exporting more to the rest of the world rather than to countries within the euro area. Since 1999, exports to the rest of the world have exceeded those to the Eurozone. The difference between the two numbers also increased slowly over time. Other peripheral countries, such as Portugal and Spain, experienced similar struggles as Greece, and thus, part of the problem can certainly be attributed to the SEM. However, this is not completely the fault of the SEM given that Greece also has difficulty isolating comparative advantages to exploit. The economic structure of Greece is inept, making it hard to thrive under a single market system because its economy relies strongly on tourism and shipping, producing little else. According to research conducted by Magoulios and Athianos of the Technological Education Institute of Central Macedonia, Greece, and published by the South-Eastern Europe Journal of Economics, Greece's comparative advantages are mostly concentrated in "labour intensive industries, low and intermediate technology, while the country still lags in dynamic high-tech industries" (Magoulios and Athianos, 2013, p. 206). These sectors naturally generate less profit and as established earlier, Greece has trouble attracting labour at a low enough cost to be economically sustainable. Lastly, Greece does not produce much and thus relies heavily on goods and services imported from other countries. Yet, according

to Magoulios and Athianos, within the Eurozone, Greece ranked fourteenth out of sixteen Eurozone states in their degree of openness, or trade percentage to GDP. Even if there are unfavourable balances of trade, the more economical import option would be within the Eurozone, given the lowered tariffs. An explanation for this trend, however, could be that given its more extensive exports with countries outside of the Union, as demonstrated by Figure 10, Greece has also negotiated trade deals which benefit them outside of the European Union.

The second issue of the Greek deficit is the excessive borrowing of money by the government. To join the EMU, Greece had already embarked on costly deficit spending programs, straining to meet the criteria set. However, upon joining the EMU by 2001, the deficit as a percentage of the GDP began to steadily increase. Figure 11 shows this obvious trend. In 2000, the deficit was at -4 percent and by 2009, it capped at -15 percent. The correlation suggests the EMU has a direct link to the deficit because of the access to capital under the monetary union. The stability of the EMU, which was anchored by several of the world's top ten leading economies, particularly Germany, France, and indirectly backed by the UK within the EU, secured the credibility of the euro, which was compounded by its substantial economic capacity. When the GDP rates of each member of the EU were added together, the total exceeded that of the United States. Furthermore, policy targets implemented by the EU, including the Convergence Criteria, also ensured the investor class's confidence in the new currency. These factors, when combined, led to a dramatic lowering of the cost of capital, and thus allowed easy access to the pool of money that could be loaned. Greece utilized this capital to cover the cost of current debts as a result of its trade deficits; yet this capital was not earmarked for programs that would provide long term solutions by generating sufficient streams of revenue to offset future debt (Congressional Research Service, Nelson, Belkin, and Mix 2011, p. 3). This explains the perpetual accumulation of deficit, and as a result, the debt accumulated. However, the ECB and the EMU failed to implement sufficient regulations by placing too much trust in governmental fiscal responsibility and favouring maximizing economic growth over stability.

While the fault of an individual country cannot necessarily be attributed to the institution as a whole, the cost definitely spills over to the entirety of the EMU. The chief problem that the EMU suffered from in the decade prior to the Eurozone crisis was its lenient approach in the Stability and Growth Pact (SGP). At the time of its creation, the primary concern that the institution had was to establish a sturdy foundation for its currency, rooted in both the strength of the German/French economies and the expansion of smaller economies. The more recently passed Fiscal Compact resolves that problem by instituting accountability for reckless economic behaviour of the signatories, which include all the EMU members and a few additional EU countries which have chosen to ratify the treaty. It steps into the bounds of national sovereignty in Article 3(2) with certain fiscal requirements, many of which already existed in the SGP, be added as legally binding provisions of the SGP. Article 8 of the Fiscal Compact is the whip preventing the non-compliance that riddled the SGP. It allows the European Commission to bring a contracting party to the European Union Court of Justice (ECJ) if the country has failed to comply with Article 3(2). This power is also granted to other contracting parties. The decision of the ECJ will be binding and the Court will grant the country a specific amount of time to take the necessary measures. If the country continues to fail to abide by the ECJ decisions, monetary penalties may be imposed on the disobedient state (Fabbrini, 2013, p. 8).

Under a monetary union that allows for different fiscal policies, downturns could lead to an asymmetric economic shock. This situation creates varying economic responses in different places that have to be addressed differently. This makes addressing a shock

particularly difficult. Figure 12 depicts the shock distinctly. The graph begins at 2007, the peak of the real GDP of the selected countries before it declined. By the eighth quarter, which occurred in 2009, the start of the Greek sovereign debt crisis, there is a clear divergence in the real GDP of the countries. This is an example of the difference in response a monetary union receives after an asymmetric shock. While the rest of Europe recovered quickly, returning to pre-crisis rates within a few years, the sum of the peripheral countries'[56] recovery rates fell far behind the stronger economies of the Union. Furthermore, the figure is measured in changes in real GDP, and thus, the values of the currencies are inflation-adjusted. To address this problem, the ECB cannot depreciate the value of the Euro to help Greece pay back its debt because this would impede the growth of larger economies. Similarly, the price cannot be raised too high or else Greece would default on its debt (Shambaugh, 2012, p. 179). These are the types of problems that are created by the economic shock. It forces the need for a bailout whereby the funds are only sufficient to prevent Greece's economy from tanking. However, an economic restructuring is required as a long-term solution. Furthermore, the bailout was obviously aimed at preventing Greece's economy from dragging the Eurozone down as it scarcely benefited the people of Greece. Figure 13 demonstrates this through a comparative graph. Looking at the red line, which presents Greek per capita relative to the EU average each year, its rate of growth initially matched that of the EU average, and at times even went beyond. However, simultaneous with the introduction of the bailout, the rates took an unwavering drop, immediately sinking to 74 percent of the EU average, a consequence of austerity measures and the lowering of minimum wage.

The cause of the deficit can be pointed at Greece's insufficient use of the European market to its advantage and also the lack of regulation within the institutions of the European Union, which allowed for the unchallenged development of the Greek sovereign debt crisis. When the world entered a recession, Greece's approach to building its economy also toppled, throwing the country into the debt crisis. Yet, the structure of the EMU made it difficult to help rescue Greece from its economic plight quickly enough. The single currency tied several countries' economies together and thereby prevented the possibility of a focused approach, such as quick depreciation of currency without consequences elsewhere with different economic conditions. Thus, the solutions put forth via costly bailouts could only be designed to simply contain the crisis from spilling over, at the cost of the Greek people.

The Eurozone crisis has proven repeatedly that the bailouts yield only limited benefits, particularly if the country in need of assistance fails to find measures to improve its economy within the timeframe of the financial assistance. The payment also stymies growth by siphoning necessary funds away from a desperate economy. Therefore, the crisis solution should not depend on the capacity of an already crippled economy to rebound. Rather, it should be up to more capable countries. A possible approach could be to have larger economies with account surpluses and below-average deficits stimulate their domestic demand, particularly for foreign goods. This can be done through the reduction of payroll or income taxes, or better yet, through offering deductions based on consumption. To ensure that the weaker economies benefit, a possible approach could be to offer rebates on specific goods that are provided by those countries. The more purchases of a type of goods or service, the more deductions can be received, until a threshold is met. These tax rebates are deductions that should be regressive in nature. Essentially, the poorer the individual, the

[56]Those countries with weaker economies in the Eurozone are known as PIIGS, an acronym for Portugal, Italy, Ireland, Greece, and Spain

higher the threshold would be. This provision is aimed primarily at the middle-class. The notion that the wealthy who have more money would contribute more to the economy is poorly founded. Asher Edelman commented on this misconception in an interview on CNBC. He identified the importance of understanding the "velocity of money", which is how much money gets spent and turned around to re-enter the economy. When the money reaches the top, it stops. The rich only spend a small portion of their money while the middle-class total expenditure will be more substantial, given the larger consumer base. Thus, offering tax rebates to minority elites will foster little increase in consumption, while offering deductions to the largest sector of the consumer base can spur the economy. While public investments on schooling and higher education can also achieve similar goals, they would require time, as opposed to immediate invigoration (Busch, 2012, p. 35).

This sort of money-financed tax cut is an alternative to quantitative easing (QE) during economic downturns. Furthermore, rather than have the money given to corporations to buy back bonds, as is the case of QE, this money would be provided to the poor and middle-class. As Ben Bernanke, the former Chairman of the Federal Reserve, stated, this concept of having the central bank or the government give the people money to consume more, thereby increasing demand, is equivalent to Milton Friedman's "helicopter drop of money" (Bernanke, 2002). In doing so, the countries with less economic capacity such as Greece may experience gains from more exports being purchased in areas of high-demand. This in turn relieves the larger economies from the burden of having to continuously provide financial assistance packages. (Busch, 2012, p. 36). In addition, this injection of money into people's hands will more than likely garner support from the people. The primary concern with this approach is that it would step into the bounds of national sovereignty. Yet desperate times may call for such drastic measures, just as the Fiscal Compact demands. Furthermore, if larger economies such as Germany would like to continue depending on a Eurozone and single market, other member countries of the institution must also be in favourable conditions to maximize benefits and increase export sales from both sides. Through this method, countries in economic crisis may experience similar recovery rates within a few years as stronger economies such as Greece or France have seen. Larger economies can be compelled to adhere to these fiscal policies by the decree of a decision-making body, assigned by the terms of the Fiscal Compact. They may vote to invoke this clause after a threshold number of countries' economies have exceeded the fiscal compact's definition of a 'balanced budget' without foreseeable recoveries within a few years.

III. Political Interactions of the Greek Government

In spite of Greece's economic situation wearing down the recovery of other Eurozone countries, it would be unlikely for the EU and its institutions to support any form of Greek exit or 'Grexit.' Figure 1 presents extensive research conducted by Eurobarometer in determining public opinions within the EU. The responses were conducted via surveys and categorized into five answers. Noticeably, the survey was asking about negative responses in regard to the Union to determine dissent against the institution. From autumn of 2003 to spring of 2009, most of the rates stay relatively constant. The change, however, is observable for all, beginning between autumn 2009 and spring 2010, or around December 2009, when the Greek sovereign debt crisis struck the Eurozone. The significance of this data is that it shows that the resistance to integration was quite strongly correlated to the Greek situation. The lack of trust for the EU was particularly disturbing, reaching 80% by autumn 2012, which was the year of the second bailout for Greece. Evidently, there had been strong discontent regarding the EU as a result of the Greek sovereign debt crisis;

therefore, it would seem rational for the EU members to begin pushing Greece out. While in the spirit of integration there are no current methods of forcibly removing a country, new legislation could have been easily drafted and passed, given the likely public support it would have garnered.

As repeatedly stated before, a primary function of a monetary union is its political purpose or the force of unity. Greece's importance in strengthening the position of the EU is neither its economic capacities nor military prowess. Instead, Greece's geopolitical importance defines its place in the Union. Greece is Europe's gateway out of the continent—to the Middle-East, Central Asia, Northern Africa, and to some extent, East Asia and Russia. In the 1980s, the European Communities welcomed Greece, albeit with some concern, with open arms for the same purpose. Despite its economic instability, high inflation rates, and emergence from the military junta, Greece, along with the other southern European countries, was the key to securing those areas from absorption into the Communist sphere of influence. Political stability in the region also comes into play regarding the transport of energy sources from the Middle-East into Europe. Furthermore, if crisis erupts on the border of Europe, Greece would be the first line of defence and bear the responsibility of handling refugee influxes. This has become increasingly important following the Schengen Agreement, which abolished border checks within the Union (Lesser, Larrabee, Zanini, and Vlachos-Dengler, 2001, p. 11). In the post-Cold War world, Greece's geopolitical standing has actually grown in importance. The fall of the Soviet Union and the Communist-bloc left behind an ambiguous separation of East and West. The unique relationships Greece shares with Russia, Iran, China, and other countries around the regions that have traditionally been the largest challengers to European power grant it leverage over the EU. Greece shares a common history with both the West, in terms of political structure, and the East, with a cultural overlap other European countries cannot parallel. In the case of Russia, Greece also shares a religious background— Eastern Orthodox Church. The Cold War had the blessing of a defined enemy from which the source of the majority of their problems on the international platform was expected and could be attributed to. In the twenty-first century, as the world attempted to adapt to a new narrative, Europe had to prepare for the consequences (Rasku, 2007, p. 13). These consequences came in the form of the unfolding conflicts in Africa and the Middle-East, where instability defined almost every regime that came to power. The recent influx of refugees is just one example of a possible crisis spilt over from Europe's neighbours and into the continent. Furthermore, the complicated relationship between the EU and Turkey, a country that includes the actual geographic borders of both Europe and Asia, necessitates a bordering ally: Greece. With the twenty-first century only just beginning to reveal pieces of itself, Greece's geopolitical importance cannot be overstated. If the EU forced a Grexit, it would almost definitely result in turmoil and disaster.

However, with regards to the Greek situation, the political disagreement is less between Greece and foreign institutions. It is much more an issue within the country itself. The system of government Greece subscribes to is also known as proportional representation (PR). Under PR, parties assume a number of seats that is proportional to the number of votes they receive during elections, usually above a specific threshold. The legislative power of Greece is vested in both the Hellenic Parliament and the government of Greece, which is headed by the Prime Minister. Usually, two larger parties hold the most seats in the Parliament, creating the main and major opposition parties. However, smaller minorities also occupy a significant of seats, resulting in the formation of coalitions that are the only way proposed policies can be passed by any one side. Prior to the crisis, the dominating governmental parties were the centre-right ND and the left-leaning PASOK. An

additional unique aspect of Greece is its population distribution. According to the Hellenic Statistical Authority, around one-third of the country's population is concentrated within the Attica (Athens) region, reaching 35.34 percent by 2011. This creates a distinct feature in Greek politics. As Dr. Visvizi aptly identified in her paper *The June 17 Elections in Greece: Domestic and European Implications*, the result of such a demographic structure is that much of the political life takes place within the capital. As such, the political decisions being made both domestically and abroad resonate very strongly in Athens, translating to heavy scrutiny on the government. This leads to a situation which demands that much of the attention of political decisions deal with the concerns of the people (Visvizi, 2012).

Therefore, to compete for constituents, the government has focused mostly on the welfare of the public sector, as it expanded significantly more than the private sector. According to Michael Heise, a Chief Economist of the Allianz SE, by 2009, one in five people of employable age were dependent on state salary, essentially making it a public sector bubble. Indeed, the approach has been rooted in Greece's modern history, and was therefore difficult to revert. From 1970 to 2009, the annual growth rate of public employment was 4 percent, totalling at a fivefold increase, while the private sector only increased by 27 percent as a whole in the same four decades (Sfakianakis, 2012). Therefore, the approaches that could be employed by countries such as Spain or Portugal during economic downturn could not apply for the already overexploited Greece. The usual mechanism governments would utilize is pumping government deficit spending into the public sector to reinvigorate the economy. However, it would then seem that government austerity measures of cutting government spending would be the easiest and most logical solution. Yet, the government's approach and refusal, or perhaps inability, to shift away from the public sector and encourage growth of the private sector impeded its success, the specifics of which will be discussed in the following section. A difficulty of communication and policy formation occurred in the post-crisis years as the PR structure of the government came head-to-head with multiple opposing parties and coalitions. The crisis that ensued in 2009 was a result of Standard & Poor's downgrade of Greek sovereign debt that pushed Greece out of the international market. This downgrade came after Greek Prime Minister George Papandreou revealed the true extent of Greece's debt. His revelations resulted in the quick decline in popularity of the PASOK government, evidenced by the rapid 44 percent drop of popular support PASOK received in 2009 to a meagre 12.28 percent that significantly diminished the party's momentum in 2012 and eventually forced it to create the Democratic Coalition (Visvizi, 2012).

The crisis also gave rise to the radical left-wing SYRIZA, which appealed to young voters. Again, the parties continued to focus on the traditional constituent from the public sector, and thus, no party reaped sufficient seats to take control of the parliament. They also refused to work with the SYRIZA government that later seized a comparative majority. A possible reason for this is SYRIZA's decision to play on their position in the Eurozone, assuming the role of a confrontational character in front of its young voters against the Union, while professing pro-European sentiment when speaking to the foreign media. The party's inconsistency held up for the Union until the government decided to hold a referendum after the institution reached out with a third bailout. There was confusion in the Greek political scene, which also hurt Greece's credibility, thereby sparking anti-Greece responses across the EU (Visvizi, 2012). Furthermore, SYRIZA denounced the reforms that would lead to increased exports (Doxiadis, 2015). While according to Pew Research only one-third of Greece holds a positive view of the EU, which feeds SYRIZA's rise, the same study showed that 69 percent of Greeks want to keep the Euro and not return to the drachma.

This gridlock barricaded sufficient structural reform that could have generated new streams of revenue to pay off government debt in the future.

In conclusion, while geopolitics has long defined the role of Greece within Europe, coming into the twenty-first century its importance has only grown in an ever more globalized world. Thus, Greece holds an irremovable seat within the EU and its institutions, unless it chooses to leave as a result of a public uproar, which is unlikely. However, its geopolitical position has also come at a cost. Through all this time, from the Cold War into the era of the European Union, and in the midst of the crisis, Greece has upheld a bilateral arms race with its neighbour, Turkey, directing substantial amounts of money to military spending. As per the agreement of the financial assistance package from the EU, Greece agreed to reduce its military expenditure to below 400 million euro. Yet that number still accounts for 4% of the national GDP. Nevertheless, it is the political scene within Greece that is the most tumultuous and uncertain, rather than its relations with the world. Political infighting has barred the government in Greece from taking sufficient action against the exacerbating economic crisis in fear of losing political support.

IV. Economic Activities of the Greek Government

An important aspect of Greece that must also be taken into consideration is its economic activities, which are inevitably intertwined with the government's political decisions to build upon the EU's reason for preventing a Grexit, from an economic standpoint, Greece is also not incentivized to leave the EMU. Greece outside of the Union would be a failed state. With their own currency, inflation would once again ravage the country, and securing favourable imports would be an insurmountable task. In addition to imports of essential goods and services, the country's source of energy must also be taken into account. As such, it is unsurprising that this same reasoning correlate with the Pew Research study mentioned above, showing the majority of Greece simply does not wish to leave the EU.

Greece's response to the economic crisis as a member of the EU, however, has been to accept the bailouts and implement austerity measures. There bailouts are the only options that both sides can agree upon, given that the ultimate goal is to prevent a default on debt. Therefore, the monetary value of the currency must be kept stable as a result of the integrated economic model. The austerity measures that followed seemed to help create a feasible addition to the bailout because logically, the response to a bubble in the public sector would be to cut government spending and increase its budget (Heise, 2015). Yet, these policies did not yield the expected results. This was the consequence of Greece's approach. Austerity's aim is to achieve fiscal consolidation, and to do so, the government could either reduce its spending or increase taxation. Greece chose to increase its budget by raising taxes rather than reduce spending, in an attempt to keep hold of the parties' voting blocs. Statistics provided by the OECD in 2014, already half a decade into the crisis, depict this trend. The portion of government revenue from personal income taxes levelled at around 21 percent, while social security contributions exceeded 32 percent (OECD, 2014). With this tax structure, almost one-third of the government's revenue from its taxes is reserved for welfare, which means the money are simply redistributed back among the population and not being invested into economic productivity. While reflecting the government's sense of responsibility, this pattern has made Greece unable to make up for the cost, evident from its slow economic improvement. Figure 11 shows the percentage of the government's budget to the country's GDP. For the past twenty years, it has remained negative, rarely able to surpass zero and not experience loss. After each bailout, one in 2010 and the other in 2012, there was a decrease in deficit. However, after a short period of

improvement, there was an evident turn for the worse once again. Figure 14 also shows that while there was an initial decrease of corporate tax rates following the crisis, it ended in 2012, and the rates once again increased, reaching 29 percent by 2015. These increases discouraged the necessary economic growth that came with the promotion of investments and business activity. These high rates would be beneficial during times of economic prosperity and prevent abusive, risky investments. With a faltering economy, however, this approach could be harmful.

Incidentally, in May 2012, with the emergence of a new government, New Democracy and PASOK forged a temporary coalition, stating "the general aim is no more cuts to salaries and pensions, no more taxes" (Sfakianakis, 2012). Fittingly, when the government decided it would stop slashing the public sector, there was a tax increase in the private sector to make up for the loss. This statement also came after the crushing defeat of PASOK. These policies have focused on protecting the electorate, which explains the lack of budgetary cost in the public sector and an increase in taxes for the private sector. However, the nature of a large public sector means the private sector would have to carry the weight of taxes in paying for the salaries of the former. By 2012, for every seven employees of the private sector laid off, only one suffered the same fate in the public sector. However, with almost 25 percent of the population unemployed at the time and the rate rising (Figure 15), the taxation increases that Greece imposed in an attempt to fund its budget became a heavy burden. Its incompatible austerity measures cancelled out the revenue it attempted to generate as a result of favouring the public sector too much.

Greece's economic policies strongly reflected the political dilemma with which the country was struggling. Neither leaving the European Union as a whole nor simply the EMU was an option, given the economic difficulties it would throw itself into without the backing of the institution, and this was realized by their own people. Nevertheless, its primary economic response to the crisis as a country within the EU has been austerity measures which have brought little in return because they were designed to prevent provocation from the people and to ensure continued political support. Furthermore, austerity measures are, by design, short term solutions to ameliorate a crisis.

Therefore, in terms of the build-up of the Greek deficit, the duality of the economic aspect of the issue—trade deficit and excessive borrowing—can be directed to two different sources of the cause. Greece's loss in terms of its trade is due to the country's lack of interactions within the Union itself, compared to its other members. However, an argument could be made that their lack of comparative advantage makes it difficult to thrive under the system. Nevertheless, being heavily dependent on import, Greece would experience a lower deficit by choosing the more economical option under the free market system. As opposed to other countries that take Keynesian approaches during economic downturns, deficit spending is no longer an option that Greece can embark on, having abused this same approach during times of economic growth. In addition, Greece's position as a member of the EMU leaves their monetary policy out of their control. Yet, the ECB, which does have control, could not implement policies benefitting Greece because consequences would reverberate across its 18 other countries. Thus, the appropriate action for the Greek government to take now is to shift their economic model. Rather than reduce budget loss with austerity measures that burden the private sector, which has proven of limited impact, the Greek government should expand its export capacities. The SEM was established to benefit the trade of all countries within the European Union and to create a supranational economic entity that could dominate the international market. The ease with which goods, services, and people can move through the Union supposedly allows it to efficiently produce products that grant the countries leverage against those outside the EU. Greece's society of

small businesses may need to encourage larger corporations to rise by granting governmental subsidies to ensure they are able to overcome labour costs within a system with a single market and single currency. Fear of losing political support should be answered with regulations on these corporations in order to prevent exploitation. With the high unemployment rate, larger businesses can ensure more employment opportunities or at least cover the displacement of jobs they would cause. Furthermore, while Greece certainly has comparatively less economic impact than many of the larger countries in the EU, they do have comparative advantages to exploit. There is great potential to expand the shipping and pharmaceutical industries. Aside from subsidies, other mechanisms can be employed to improve those businesses such as infrastructural spending on seaports for increasing the capacity of the freighting industry and on promoting tourism (Busch, 2012). To look even further into sustainable growth, as a developed country, Greece is bound to see a decelerating economy. Foreign investments for the fast-growing economies of the world will be able to ensure a share of their successes and, therefore, should be necessary (Doxiadis, 2015). Only through these means can Greece break out of its endless cycles of bailouts and finite austerity measures which have failed to provide an economically-reasonable solution.

V. Sources of Backlash- Social Effects for Greece and A Demographic Breakdown

Aside from the analysis of the Greek government, it is important to note that the economic crisis impacted the people hardest. One way to determine the social effect is to look at the Human Development Index (HDI). The HDI is not necessarily a reliable source, however, because of the short time frame of the issue. Indeed, on the HDI, Greece is still ranked high in the list at 29. Furthermore, the HDI is composite in nature, comprised of life expectancy at birth, expected years of schooling, and income per capita. While the latter may have decreased dramatically, the other two would barely have changed. In fact, during an economic downturn, years of schooling may increase as the youth population seeks to secure a better education before entering the workforce. Overall, the change is insignificant, and therefore the HDI is not an accurate measurement of social impacts.

The focus, instead, should be on individual aspects of social effects such as the unemployment rate, poverty rate, and level of homelessness. As established previously with Figure 15, the unemployment rates have seen a substantial increase, jumping from 7 percent to around 25 percent since 2012. However, the significance of high unemployment is the result of a substantial portion of the population unable to generate sufficient income. Following suit is the gradual increase of poverty. Predictably, by 2014, Eurostat reported a high rate of 36.0 percent of people in Greece who are "at risk of poverty or social exclusion" (AROPE) in 2014, shown in the appendix by Figure 16. AROPE has a broader definition of poverty rates, comprised of three factors explained below, which also have considerable social impacts that may spur opposition from the people lacking access to certain living requirements.

First is the measurement for 'at-risk-of-poverty.' As defined by Eurostat, this rate is calculated as the proportion of people who have equivalised disposable income below the threshold of 60 percent of the national median. To further elaborate, the equivalised disposable income is the total amount of income a household has left available for spending or saving, divided by the number of household members that are converted into a standard set of 'equalised adult' delineated by Eurostat (Eurostat, 2014). This aspect is important given that it takes into consideration the low income in comparison to the remainder of the country, allowing for a relative analysis by an individual-country basis. While this does not always imply a low standard of living, with regards to Greece's low household disposable

income average of USD $18,099 a year, the rates would not be particularly encouraging (OECD, 2016). This number is according to OECD statistics which has an average income of USD $29,016. Given that the 'median' is a variable used, the poverty level is relative to the current circumstances, so an increase in the rates is a perpetuating drop for the economy.

The second factor is a measure of 'material deprivation,' a state of economic strain whereby an individual is unable to afford certain assets or costs for a desired or adequate life. The evaluation is based on the percentage of people unable to afford three of nine items listed by the Social protection committee of the European Commission. The 'criteria' range from payment of rent or mortgage to more luxurious and unnecessary items that some deem necessary, such as a television set or a car (Eurostat, 2016). Nevertheless, he statistics are not determined by the lack of any one of these items, but rather, the inability to pay for one. Furthermore, the capacity to pay would require consideration of the cost of those 'materials' and, therefore, would also be dependent on specific countries, a reflection of the PPP.

The final factor is the indicator of 'persons living in households with low work intensity,' which is the number of people within a household who have a measurement of work intensity below the threshold of 0.2. Work intensity is defined as a ratio of working-age members of a household and the total number of months they have spent working in any given year divided by the total number of months they had the capacity to work. The criteria for a working age person would be one aged 18 to 59 with exceptions for those who are still in school from 18 to 24 (Eurostat, 2014). Unlike the two other standards, a measurement of work intensity looks beyond income. In an economic downturn, rather than simply dropping wages, high numbers of persons living in households with low work intensity imply a lack of job opportunities and, therefore, probably high unemployment rates. A possible criticism for this metric is its lack of consideration of work patterns within individual countries. Certain countries may demand less time for earning similar wages or vice versa. However, a ratio of 0.2 is low for any country and most jobs require a longer work time to earn adequate wages; therefore, this factor still accurately contributes data.

While AROPE offers a relative analysis of poverty rates within a given country, it lacks a certain comparative aspect that could prove to be a determining factor for discontent; this important metric is the relative economic well-being in the context of the country's living standard over time. OECD introduces a measurement of this with 'anchored poverty,' which focuses on "looking at how [low earners] are doing now compared to sunnier economic times (OECD, 2014)." Eurostat employed the same method on their AROPE data in Figure 17, with the reference date anchored at 2008. Indeed, the reported anchored poverty rate was over 10 percent higher in 2014 than the original AROPE rates, bordering 50 percent. Essentially, the poverty threshold is dependent on the economic circumstances of the time and the median that fluctuates as a result. Therefore, Eurostat simply utilized the threshold during the last year with a favourable economy in Greece—2008—and applied that number to the conditions in both 2013 and 2014 (Eurostat, 2015). The value of anchored poverty is its consideration for the expected economic situation that the people would be accustomed to, having lived through better conditions. *Statistics do not dictate how people respond to the reality of their position*. People who are on the edge of poverty, as defined by the original approach to poverty of a country, may already feel their existence has stooped below the threshold. This is because in comparison to better past lifestyles, their current predicament would be relatively destitute.

Associated with low poverty rates is the level of homelessness that ensues. By 2015 in Greece, rates reached as high as 40 percent increases in three months in the capitol city of Athens. Just within the urban area of Athens, and not the Attica region as a whole, the Greek government estimates there are 20,000 homeless people in a city of 660,000.

Furthermore, as previously stated, the population distribution of Greece is unique in its heavy concentration in and around the capital of Athens, where the level of homelessness has also been highest. This would catalyze conflicts against the government especially given the indignifying living conditions these people have to endure while they remain unable to afford a shelter as a result of the economy. The impacts of the economic downturn are obvious to the Greek people. High rates of poverty, unemployment, and homelessness are explanations of the lack of economic movement towards larger businesses and support of investments to spur the economy because the leadership of the government would fear losing political support. Therefore, their austerity programs focused on ensuring that the general population would not seek political change by attempting to keep as much of the public sector intact as possible.

A necessary component to determine the implications of the current crisis is the demographic breakdown of the sources of the backlash. In the case of Greece, the most obvious trend is in the plight of its youth population. Figure 18 depicts the ballooning of youth unemployment rates in Greece, reaching up to 60 percent in 2013 and levelling at 50 percent. This unfortunate number means slightly over half of all people between the ages of 15 and 24 actively seeking a job are not getting one. The dramatic increase suggests there is a lack of opportunities for an occupation in Greece as a result of the economic situation. This trend is reflected in two other social impacts discussed. Figure 16 extends its statistics with a demographic breakdown by age group. While Greece already has a bleak AROPE rate of 36 percent overall, it is the working population that suffers most from poverty, reaching 40.1 percent. Within that particular group, the young people—aged 15 to 29—can be further broken down. Young people from Greece have a slightly lower AROPE rate of 29.3 percent, largely because many would still have a family to support them in their earlier years and poverty can be temporarily averted. However, the rates are not as "favourable" for those outside the country. For foreigners in general, the rate skyrockets to 54.3 percent and for those not from EU-28 countries, poverty reaches a staggering 74.8 percent (Eurostat, 2014)[57]. A troubling implication for these astronomical figures is that they strongly discourage the movement of people into Greece, which is vital to support and, in the future, to sustain the economic growth that the usual population growth rates in developed countries simply cannot accomplish. At this scale of poverty for young people, it is unsurprising that people aged between 26 and 45 now make up the biggest group of homeless in the capital city.

All the data that outline the multiple effects of the Greek crisis point in the same direction: the youth have bore the brunt of the economic downturn. As a result of the mass deterioration of social conditions for this generation, the first response would be to question the government. While traditional parties such as PASOK continue trying to hold on to the remaining public sector, the younger population sought an alternative that could change a system that seemingly excludes their wellbeing. The PASOK leadership and their actions that led to the beginning of the economic crisis gave rise to SYRIZA in 2012 which attracted young people, mostly from ages 18 to 44, as their main support base. SYRIZA's confrontational approach promises profound changes, particularly with their position against the EU which most other parties have not supported. However, movements toward drastic social and governmental changes spawned responses from either side of the political

[57]Eurostat goes into further detail and breaks down the data into different categories covering: education, employment, and social inclusion of the youth population in Europe as a whole; for further information, see: http://ec.europa.eu/eurostat/statistics-explained/index.php/Young_people_-_migration_and_socioeconomic_situation

spectrum. By 2015, the far-right Golden Dawn had become the most popular party for people aged 18 to 24 (Sakellaiou, 2015, p.9). The surge of popularity for both radical groups, which supposedly represents polar opposite positions, from this young demographic group, suggests that the support probably does not come from specific policies proposed by either side. It was more than likely due to a desire from the youth to tear down the establishment that has clung to the centre left and the centre right—PASOK and ND, respectively. From the Golden Dawn, for example, a qualitative poll conducted on young people in the party stated they denied their leaders' fascists ideologies, recognizing them as nationalists (Sakellaiou, 2015, p.11)[58]. The continuation of PASOK and ND's collaboration with the EU has caused dissent, as overwhelmingly, 85 percent of Greece said the Union did not understand the needs of the Greek people (PIIE, Stokes, Wike, and Poushter, 2016). This is mirrored in the voting behaviour of the 2015 bailout referendum. 61 percent voted 'no' to agree with the bailout, and the vast majority of SYRIZA's voter base concurred. Yet their reasoning was not due to their unwillingness to cooperate with the European institutions. Instead, it was because they desired a new approach and did not understand the implications of their votes. They do, however, recognize that they have suffered through five years of poverty and are expressing their disapproval with the current austerity measures (Daley, 2015). With the youth population identified as the primary source of the backlash, and given that economic reforms are dependent on the ability for the government to gain political support for their policies, actions taken by any political party should be designed in a way that also ensures the approval of this agitated demographic group.

VI. Implications of the Greek Crisis

In regards to the Greek situation, the largest concern for the remaining EU countries is how Greece's economic plight would impact them, particularly those within the monetary union. The hefty bailouts have been the primary focus of contention, both within and outside of Greece. While these financial assistance packages will total hundreds of billions of Euros, the costs are arguably necessary to prevent the aftermath of the alternatives. There are two alternatives if a bailout were not provided: allowing Greece to default on its debt or pushing for a Grexit. The first option would lead to an immediate mass withdrawal of capital from Athens as the credibility of the country freefalls. This would throw Greece into an even more severe economic depression as the country is pushed out of the global financial market. Therefore, Greece's only option in a situation like this is to resort to the second option: to reintroduce the drachma in order to depreciate the value of their currency and pay back the debt. In both cases, the ECB would likely be thrown into the spotlight and questioned for its decision to not provide the necessary aid for a country in the monetary union it oversees. A similar response from the investor class would be projected on the Euro's lack of stability and inability of the ECB to keep the members of its monetary union from defaulting or resort to leaving and hyper-inflating their currency to prevent the catastrophe (PIIE, Kirkegaard, 2015).

Figure 19 proves that many of the larger economies such as Germany and France had made a full economic recovery by 2011 and the Eurozone overall has accomplished the same by 2016. In fact, given that the Eurozone data takes into consideration the low rates of Greece, excluding Greece will show a much speedier improvement of the economy for other countries. Greece is the outlier to the trend and the remaining countries do not seem to

[58] The demographic breakdown by age group regarding support for political parties is visualized in a table presented on the ESRI story map. The original details were released in an article by Sakellaiou, the details of which are covered here: http://library.fes.de/pdf-files/bueros/athen/11501.pdf

follow a similar correlation. It is the speculation about Greece's situation from the investor class that has impeded the process of the Euro rebounding. During an economic crisis, the investors convert their money to the safest option, which would be US currency. The United States' Quantitative Easing (QE) policy stabilized its economy quickly while the EU was in the midst of the Eurozone crisis, with the Greek sovereign debt crisis soon stacking on (Congressional Research Service, Nelson, Belkin, and Mix 2011, p. 14). The concern at the time was that fear of a Grexit would spur a bandwagon effect whereby countries such as Portugal, Spain, or Ireland would also leave the Union. This would cripple the Euro and render the currency uncertain regarding the ability to continue the monetary union. However, both the European Union and the leadership of the Greek government understood that a Grexit was not an option. Aside from the Communist party, the major parties realized that the even higher inflation rates and lack of economic strengths caused by a return to the drachma could not compare to the benefits reaped under the monetary union. This understanding between the political groups includes SYRIZA, which has thus far chosen to simply dangle along the lines in regards to their position on the EU while not actually crossing the border. Furthermore, Alexis Tsipras, the Greek Prime Minister, has agreed to the third bailout agreement for 86 additional billion Euros to be given to Greece over the course of three years, despite strong public opposition. Euro's depreciation was actually caused largely by the decisions of the ECB. To reach the ideal inflation rate of 2 percent, the ECB has decided to pump a trillion Euros between March 2015 and September 2016 in their own QE program. Additionally, they decided to set record-low interest rates of 0.05 percent from 10 September, 2014 to 16 March, 2016, before they were lowered to 0 percent (ECB, 2016)[59]. The obvious purpose of these policies was to increase the money supply in the system drastically, thereby decreasing the value of their currency and increasing inflation rates. This would allow for increased demand for goods and services from the Eurozone as European exports become less expensive for overseas buyers.

Using Greece as a marker for failures in other countries would be an inaccurate assumption given that Greece's economic problem is largely tied to its political situation and also due to the fact that Greece joined the Union already heavily burdened by deficit. As discussed, within the EU, Greece is a peripheral country, both geographically and economically. The flow of labour into Western Europe would be substantially larger than to countries such as Greece. Furthermore, Greece's economy depends largely on tourism and shipping, which together contribute a quarter of the GDP (OECD, 2016). Unfortunately, during an economic downturn, the tourism industry can be particularly hard hit. The number of visitors drops drastically when people are no longer able to afford these trips to visit. Alternatives are difficult to come by when tourism is among the largest industries of the Greek economy. However, certain features of Greece's economic structure could create the opposite effect. An example of this would be the shipping industry. In Europe, the two dominating shipping countries are Germany and Greece. Germany's shipping businesses rely largely on the *"Kommanditgesellschaft*" (KG), limited partnership, system which encourages numerous investors to put their money in the industry for a return on investment (Paris, 2015). Greek shipping magnates, on the other hand, kept the arrangement concentrated within their own businesses. Therefore, when the freight market continued to fall after the economic crisis, the investors in Germany were quick to pull out due to the loss of profits. This led to a domino effect, and by 2014, up to a thousand of more than three thousand ships owned by German businesses were up for sale as German banks tried to

[59]For more detail, see: https://www.ecb.europa.eu/stats/monetary/rates/html/index.en.html for specific data

leave the industry. Greek owners quickly grabbed pieces of Germany's crumbling shipping industry, purchasing high-quality ships at low prices.

While shipping and tourism contributed heavily to the GDP, by far the largest portion of Greek GDP is its government spending, reaching 60.8 percent in 2013—the highest of all OECD countries. This again reflects Greece's leniency on big government directing resources to the public sector. While high government spending does not necessarily imply deficit, in the case of Greece, high social spending without enough returns has left greater deficits than other countries such as Slovenia, France, and Belgium, the next three EMU countries with high rates of government spending. Therefore, given the unique circumstances of different countries, it would be incorrect to claim the Greek situation could be a foreshadowing of other Eurozone economies. While the shipping industry proves to have managed to find generous silver-linings to the economic crisis, Greece's economic position in other sectors has made it difficult for the country to recover as other countries did. Its reliance on tourism and the political arena's excessive focus on the public sector have led to high deficits the country simply cannot make up for. Similarly, the economic difference is not solely a matter of centre-periphery imbalances. Within PIIGS—Portugal, Ireland, Italy, Greece, and Spain—the cause of their economic problems also contrast. Ireland suffered for lack of supervision on their banking sector's excessive borrowing of money, and as a result, several of their major banks went bankrupt in the 2008 economic crisis. Their lack of money led to both a fluidity and solvency crisis. While Spain had its debt and deficit under control, it allowed for a housing bubble which burst with the 2008 crisis, leading to heavy borrowing to offset the property collapse. Lastly, high deficits caused both Portugal and Italy a debt crisis, Greece's dilemma was a result of decades of policies designed to buy political support in exchange for excessively boosting the public sector (Shambaugh, 2012, p. 161). These Eurozone countries with weaker economies have structural issues of their own that must be addressed differently in response to separate causes.

Another potentially worrying aspect of the Greek situation is the spillover of political and social opposition. As previously discussed, Figure 1 demonstrates anger directed at the European Union resulting from the Greek crisis, and its timing synchronizes with the beginning of the Greek situation in 2009. However, the data of Figure 1 fails to look at the later years when the economies of major European countries rebounded from the downturn. The Pew Research Centre conducted a series of studies that extend into 2014, suggesting improvement from the initial antagonism. Figure 20 shows that though most countries, in the beginning, lost confidence in European integration as beneficial to the economy of their own country, by 2013, they experienced a revival of tenacity in the EU project. It is unsurprising, therefore, that the changes in trend correspond to significant events during the timeline of the crisis. From 2009 onwards, there have been decreasing percentages of people who believed their country was benefitting from the integration. In 2012, the slope of the line got even steeper, signifying a faster drop in popularity. Incidentally, this was the year of the second bailout. Returning to Figure 19, the data reports that the Eurozone overall began seeing the economy improve once again, gradually, after the slow decline from late-2010, a few months after the 'First Economic Adjustment Programme of Greece.' Indeed, this correlation is supported by the Pew study. Spain, which saw little improvement, and Italy, whose economy actually continued to drop after 2013, were among the bottom few rankings in terms of changes in GDP, which remained negative after the recovery of most countries. Therefore, their lack of support could be expected since both countries were still worse off than they were in 2009. Furthermore, while the youth were identified as the demographic group suffering the most in the economic crisis, statistics showed similar trends to the

previous data. The percentage of 18 to 29 year-olds favourable to the EU all dropped from 2007 to 2013. However, from 2013 to 2014, all—except Poland which remained at 0 and Italy which decreased—youth from the listed countries reconsidered their position. This includes Greece where the change was by +8 percent. The same demographic group increased its support of integration as a means of strengthening the economy(Pew Research Centre, 2014)[60]. A plausible conclusion is that public opinion relates very strongly to the economic welfare of the country. Therefore, so long as the Greek economic situation is contained within one country, a strong social response in opposition would be unlikely to occur in other European countries.

The opposition to European integration efforts throughout Europe has consistently come from the right-wing. In the 2014 study, Pew identified the 'Left-Right Difference' in terms of 'percentage favourable for the European Parliament.' If the number of the difference was positive, then the left had more support than the right. In almost all of the cases, that number was positive. The largest difference in opinion was in the United Kingdom where it reached +17 percent, with 45 percent of the left having a favourable opinion of Parliament. While the numbers were overall not encouraging for integration, the bulk of the opposition was made up of right-leaning voters. The exceptions to this trend, however, were Greece and Spain, where the left showed less support than the right. In Greece, however, this could be attributed to the austerity measures implemented as a result of the bailout packages which angered the left.

Nevertheless, the responses in the majority of Europe demand attention be paid to the right-wing opposition to integration. Herbert Kitschelt, an International Relations professor at Duke University, identified the key characteristics of the 'new radical right' in Europe that discern it from its previous approaches. In the twentieth century, far-right parties had 'movement support' such as that seen with traditional fascism. Today, supporters of the same political spectrum lack strong structure and organized movements in their party organization. As a result, the agenda they pursue has high volatility and is quickly changed in response to immediate circumstances. This, however, also contributes to their quick surges in support as the new radical right are largely single-issue oriented (Wodak, KhosraviNik, and Mral, 2015, p. 10). Furthermore, while they certainly lack in the focus of their concerns, right-wing politics still have a common ground they hold onto and that is their reliance on *identity narratives* to garner support by creating clear delineations between an in-group and an out-group (Wodak, KhosraviNik, and Mral, 2015, p. 26). The Greek sovereign debt crisis saw an immediate increase in doubt of the benefits of being in the EU. While each member state certainly wants to reap the benefits of integration, as soon as a bailout plan was put forth, opinion quickly shifted to question each country's responsibility in carrying the weight of Greece's economic plight. When eventually the Greek situation pacified, the migration crisis ensued and the identity narratives became clearer, with the issue becoming a racial one. In the United Kingdom, there has been a particularly strong and drastic response. Support for the 'Brexit' increased dramatically since the start of the migration crisis. In 2014, the majority still supported EU membership. However, a steady increase of refugees into Southern UK closed the gaps between the two, as both sides competed to win by small margins (Pew Research Centre, 2014).

While the situations may seem exclusive, concerns about social issues cannot be separated from economic matters. With the third bailout package underway and Greece's economy scarcely improving, it would not be hard for political parties to tie the social

[60] See for specific data: http://www.pewglobal.org/2014/05/12/chapter-2-crisis-of-confidence-in-the-eu-ending/

aspect to economic burdens. Undeniably, the containment of public perception of the crisis in Greece is of great importance. Migration from the East, the United Kingdom's relationship with the EU, and Greece's still-precarious economy put pressure on the EU to ensure public opinion does not stray towards disturbing trends of disintegration.

VII. Sources of the Backlash- Lack of a 21st-century European Narrative

While economic integration is certainly an essential factor in creating a successful Union, the political aspect must also foster an environment supporting cooperation. An argument could be made that backlash could be better prevented with the establishment of a new *European narrative* or *European identity* in which the people of different institutions feel akin to[61] [see *Acknowledgments* section at the end of the paper]. Perhaps one of the best pieces of modern literature to break down the European scene is Charles A. Kupchan's *How Enemies Become Friends: The Sources of Stable Peace.* Kupchan recognizes three conditions that are present in a situation of strong and stable peace: strategic restraint, compatible social orders, and cultural commonality (Kupchan, 2010, p. 10). The essence of strategic restraint is governments' exercise of limitations on their own powers. This facilitates the process of peace because only by accepting these constraints can alliances be forged, or else competition for power would prevail. To a significant extent, the members of the EMU accomplished this when they gave over control of their monetary policy to the ECB as part of the criteria for single currency. While institutionally, those countries may certainly have accomplished restraint, Pew Research Centre has conducted polls to determine if the European people would still be willing to continue down this path. The majority, except in Poland, oppose giving more authority to the EU, reluctant to allow their country to give up power over their own policies (Pew, 2014). This would easily correlate to dissent for the Greek bailouts and backlash against the Union's decision to provide this hefty financial assistance package. The money that is already being given away from individual countries to the ECB is now being delivered to a country on the periphery, Greece, without the consent of the people, indicating a clear lack of power in the decision-making process.

The second aspect of stable peace is the presence of 'compatible social orders.' Essentially, it calls for social structures in different countries to mirror one another so the integration process does not disrupt systematic hierarchies in place (Kupchan, 2010, p. 11). If political and economic elites belong to people with very different values, the path towards peace may be obstructed. These values may come in terms of religious or ethnic background, a problem the European Union has yet to face as their memberships have remained within the small continent. This would explain, however, the difficulty Turkey is experiencing in gaining the approval of the European Union after years of trying to integrate with the West. Their religion and ethnic makeup differ significantly from those of the EU. The values may also be dependent on the political or social order. With integration and increasing interaction between peoples, the distribution of political power among social classes in particular may threaten the existing authority a different group has in another country.

The final condition is a cultural commonality or, as Kupchan described, "interlinked network of practices and symbols based primarily on ethnicity, race, and religion (Kupchan, 2010, p. 12)." These latter two goals have thus far proven to be sufficiently reached by members of the European Union. With an extensive common history and having shared

[61] This portion of the research essay focuses primarily on identifying key the problems within the EU, currently, that contributes to the difficulty in maintaining stability. Regarding potential approaches to establishing a new identity, Kalypso Nicolaidis has proposed several answers; for more information, visit: *http://arcg.is/29VMl80*

comparable, if not the same, regimes in the past, the social and political orders of Europe are certainly compatible. For these same reasons, countries in Europe have similar cultural practices and, due to centuries of exposure, the differences have yet to be a major diverging factor in recent decades. Particularly since the twentieth century, measures have been taken to improve relations between the countries and promote social contact through international sports events or cultural gatherings, such as art exhibitions. However, the migration crisis, coupled with economic stress, has become a potential wedge in the effort to establish a common European narrative. It has spurred a new test for Europe's ability to fulfill Kupchan's criteria. The willingness to accept different ethnicities varies among European countries. The United Kingdom's backlash displayed a drifting away from the European project.

These conditions play out in four separate stages to establish stable peace: unilateral accommodation, reciprocal restraint, societal integration, and development of a generation of new narratives and identities. First and foremost is *unilateral accommodation* which occurs when a state decides to respond to the existence of a powerful threat or multiple smaller ones by removingthe source of a potential threat by exercising strategic restraint and making concessions to its former adversary. The initial effort made by European countries to establish the European Coal and Steel Community and the later European Communities were for these purposes. To consolidate their fragile alliance and prevent the spread of the Communist threat, the 'Western-bloc' created these institutions to parallel their neighbour's strength. The primary rival in this case was the Soviet Union and the divisions were clear. In fact, the separation was famously branded by Winston Churchill as the 'Iron Curtain' whereby both sides exercised strategic restraint to achieve unilateral accommodation. Indeed, the European Community's eager acceptance of Greece into their institution was to secure Southern Europe from being absorbed into the Soviet sphere of influence. With the fall of the Soviet Union, the need for unilateral accommodation was largely buried under the rubble of the Eastern-bloc. While the institution may still recognize the geopolitical importance of Greece, the European people have ranked Greece the least trustworthy in a 2013 poll, tied with Italy (Pew, 2013)[62]. The EU itself tried to reinforce the practice of strategic restraint through the continuation of the Common Foreign and Security Policy (CFSP). This, however, has failed to accomplish unity in terms of inspiring mutual assistance within the EU. Most Europeans believe that their country should "deal with its own problems and let other countries deal with their own problems as best they can." In fact, the median rate of Europeans who held that view was 56 percent (PIIE, Stokes, Wike, and Poushter, 2016).

The second phase is the practice of *reciprocal restraint* whereby countries tread slowly in exercising tit-for-tat restraint in power, gradually entertaining the possibility of cooperation rather than rivalry. The European Union initiated this stage with the Treaty of Rome early in 1957 which set the framework that led to the implementation of the SEM in 1987. This leads to *social integration* whereby the transactions between countries that have begun forming an alliance increase in frequency, leading to more interaction between the governments, markets, and people. Interestingly, Kupchan recognizes this as the third stage of his approach to establish stable peace. There is a recognizable difference from the neofunctionalist approach on which Monnet based his policies and built the European Communities. Monnet believed achieving economic interdependence was a priority before political unity could follow. Kupchan, on the other hand, believes political compatibility and

[62]See for opinions of European countries about one another:
http://www.pewglobal.org/2013/05/13/the-new-sick-man-of-europe-the-european-union/

union is the starting line and economic integration is a tool used to strengthen the peace. As in the case of Greece, when one of the economies in a monetary union crumbles, certain countries take the hardest hit and others recover quickly—the symptoms of an asymmetric economic shock. Thus, while a few countries trailing behind burden the economy overall, a backlash against economic integration under the second and third phases may develop alongside worsening economic conditions. However, if there were a strong need or desire for political unity as is presumed by Kupchan's model, the Union may not fall apart and peace could be kept. Unfortunately, for the European people, that has yet to be the case.

The final phase is the development of a *generation of new narratives and identities.* Essentially, through all these processes and stages, there would be an emergence of a shared sense of solidarity, culminating in a singular identity (Kupchan, 2010, p. 10). The creation of the EMU not only bolstered the economic strength of Europe, it also created a new political purpose for the European Union in the post-Cold War world. Under a single currency, there was hope for a common identity based around economic interdependence. Therefore, unsurprisingly, the success of this new political alliance is heavily dependent on the economic success of the continent, given that the people would draw a direct link from economic downturns to the EMU. In such a situation, preserving and ensuring the success of the EMU is necessary for the economic prosperity of its members and also for the survival of a stumbling political narrative. As of 2014, the Euro has certainly retained strong support from the European people. In Germany and France, the chief contributors to the bailouts, 72 and 64 percent, respectively, of their people want to keep the Euro. In both Greece and Spain, where the Eurozone crisis hit the hardest, support still remained at 69 and 68 percent, respectively (Pew Research Centre, 2014). Nevertheless, the Eurozone has failed to convince the inclusion of all the EU members. Furthermore, polls conducted by Eurobarometer have shown that the Euro has failed to affect people's opinion on a *European identity.* In a brief published in September 2009, 76 percent of the people who live in EMU member countries reported they felt the Euro had no influence in creating a common identity. Only 23 percent felt "more European" as a result of the single currency. In 2007 and 2008, prior to the Eurozone crisis in late-2009, the rates were also very similar: 77 percent did not change their opinion (Eurobarometer, 2009 p. 9). This data is reflected in 2010 and 2015, as well, following the start of the Greek sovereign debt crisis, staying at 77 and 72 percent, respectively (Eurobarometer, 2010, p. 6; 2015, p.15).

While the Euro may have failed to create a common *European identity*, perhaps a stronger aspect of the EU that may accomplish this is the Single European Market. The free movement of goods, services, and more importantly, people, will continue to increase interactions among peoples of different countries. Yet the effects of this have been limited. Establishing a new regional narrative requires people to stop attributing their identity solely to their countries of birth or residence. This process demands much more time than the thirty years since the SEM started; preventing an economic downturn from hindering this process is imperative. Nevertheless, thus far economic integration has not necessarily proven effective at creating a *European identity* to prevent backlash from the people of the member countries of the institution, given that its dependence on economic success can be undermined by a situation such as that of the Greek sovereign debt crisis.

Conclusion

This research began with a case study of the Greek integration into the EU—its efforts to join the Union, the economic burdens which were demanded, and its behaviours that followed membership. With the scene set, the analysis delved deeper into the individual causes of the Greek sovereign debt crisis: Greece's irresponsibility within the EU and its

structural imbalances, the turmoil in the Greek political arena, and its economic activities that drove the country into heavy deficit. While separately explained in different sections, the three certainly were not mutually exclusive. Rather, they were intertwined, each a synergetic component to the build-up towards debt crisis. Arguably the underlying cause of the Greek plight, however, has been political infighting which has led to unsustainable economic policies, particularly the ballooning of the public sector in the face of tremendous deficit burdens. This translated to the rapidly shrinking portion of the Greek population employed within the private sector and, as a result, thwarting the Greek government's efforts at generating sufficient tax revenue. In addition, being within the EU has forced Greece to compete with countries it cannot match in economic strength, cutting into the government budget, further driving up trade deficit. Therefore, following the global recession and the sudden sovereign debt crisis from the Papandreou administration's admission about the actual extent of the Greek financial situation, the government was overburdened with paying off its own debts. All the while, they were attempting to continue their approach and sustain payment of salaries to those in the public sector, desperate to keep their voting blocs intact. These trends have been reflected across the myriad of data. Yet in all of the cases, certain measures can definitely be taken to reinvigorate the Greek economy and restore its stability. While the EU needs to work on institutional changes and more effective responses to economic crises within their union, Greece needs to shift from a dependence on the public sector.

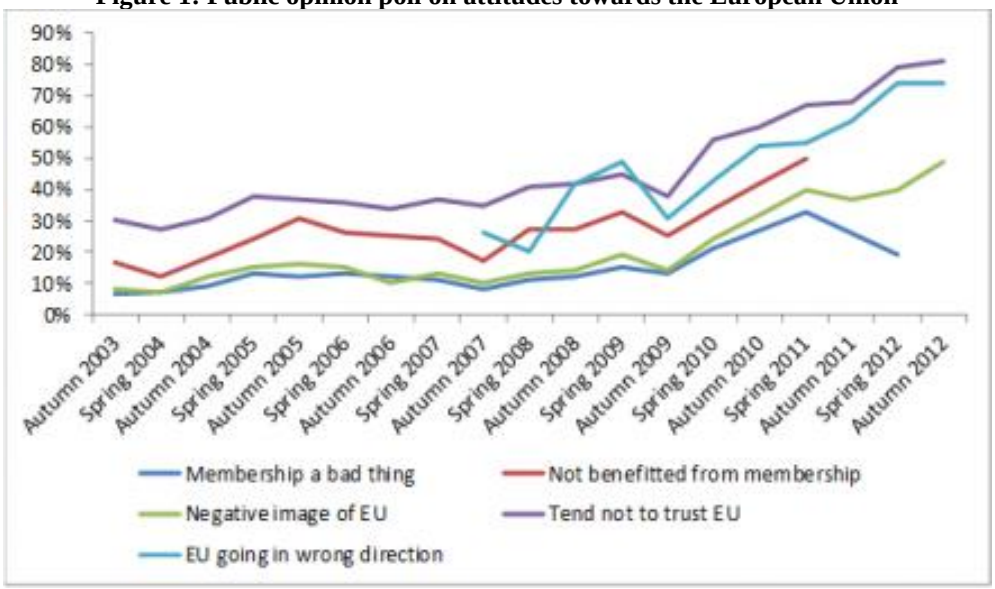
The significance of the sovereign debt crisis, however, is represented in the consequences that ensued. The economic impacts were clear as Greece's GDP dropped by over 25 percent. The effects on Greek politics were also apparent. The establishment saw rapid declines in support, and younger, more radical parties, especially SYRIZA, experienced rapid growth. The demographic breakdown dissected the backlash of the debt crisis and offered an explanation to the responses. Without a doubt, the Greek people suffered a striking loss in living standards. Unemployment quickly increased, as did the AROPE rates, a measurement of poverty and level of homelessness. Nevertheless, these events are largely sequential by nature, and, therefore, unsurprising. To determine the true extent of the social impacts imposed by the debt crisis, OECD's 'anchored poverty' was employed. Taking into consideration the standards of living a few years prior by setting the poverty threshold to the number determined on the 'anchored year', the decreases were much more dramatic. Overall, the Greek people took a major hit from this economic slump. The age group most affected were young people, who experienced unemployment rates of beyond 50 percent, explaining the quick rise in support of the youth's primary choice of party: SYRIZA.

With the Greek narrative set within the context of the EU, the implications that the Greek situation poses for the Union as a whole also bear importance. Many fear that the economic effects experienced in Greece will, to some extent, be imprinted on other countries, belonging to the same monetary union. Yet the consequences of the Greek crisis have proven to be of limited significance. The misattributed drop of the Euro was actually a result of the ECB's economic policies–its attempt to jumpstart the economies of EU countries through increased exports. Furthermore, Greece's geopolitical importance to the European continent as well as the country's dependence on the Union itself has created a situation whereby Grexit is simply not feasible, despite the stance SYRIZA has often purported. Finally, by applying the overarching European situation to the publications of Charles A. Kupchan, the EU and its monetary union project has certainly found it difficult to accomplish its goal of establishing a new European narrative.

This research has demonstrated that the Greek situation was the product of both the flaws in the EU and its institutions, as well as Greece's political turmoil which allowed for economic imprudence and unsustainable policies. Also, in spite of the slow improvement, Greece's situation is a necessary cost that, while burdensome in the short-term, is the far better alternative than to refuse to help the faltering country. With correct measures that can be taken by both Greece—in its restructuring of its economic model to support the private sector in exports—and the EU—to create more stringent policies and more effective crises responses—the Greek situation can be resolved and averted in the future. However, the difficulty of convincing the people of the EU without stirring opposition indicates a lack of political unity within the Union that is dependent on economic prosperity. Therefore, the question that now remains to be answered is this: As the Euro has failed to create and promote a common identity among the European people, how will the Union now replace the single currency in order to establish a political union that goes beyond dependence on economic prosperity, and rather forges a new *European narrative*?

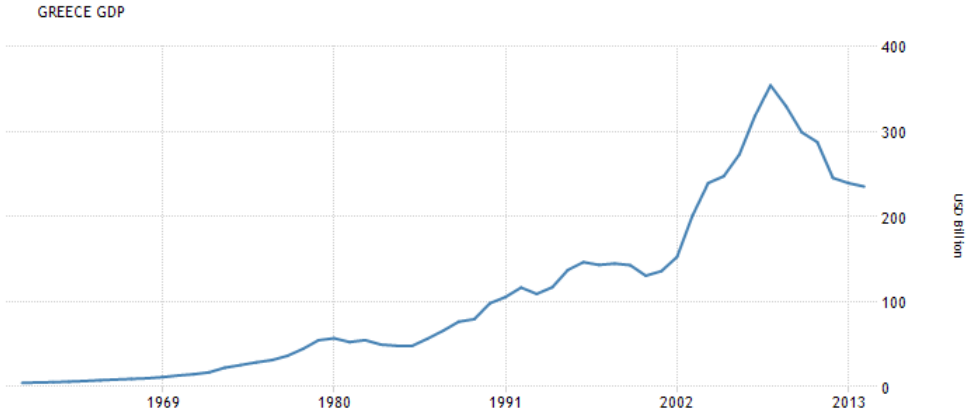
Appendix

Figure 1: Public opinion poll on attitudes towards the European Union



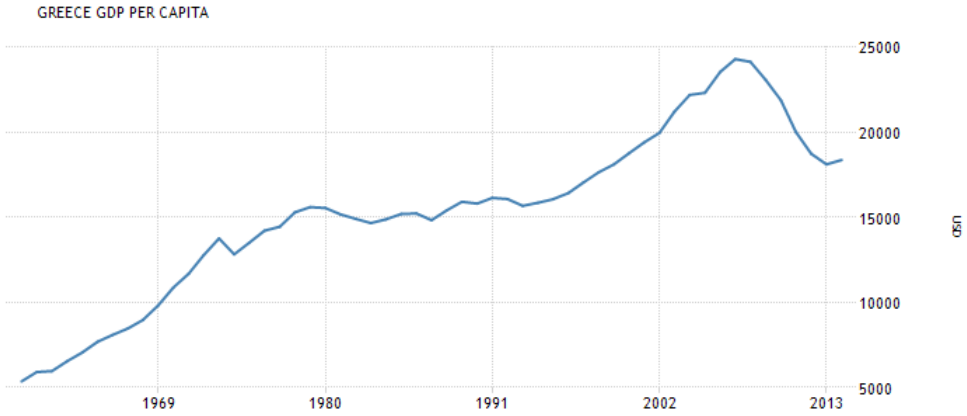
Source: Eurobarometer Surveys

Figure 2: Greece GDP from 1960-2014



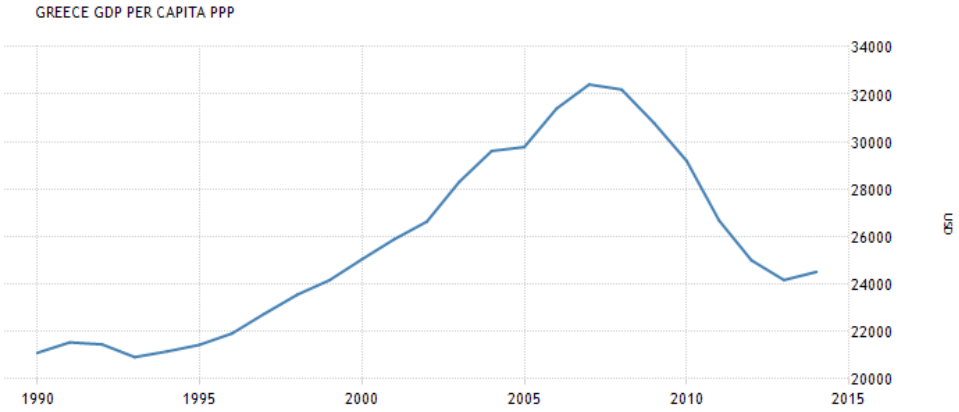
Source: tradingeconomics.com | World Bank

Figure 3: Greece GDP per capita from 1960-2014



Source: tradingeconomics.com | World Bank

Figure 4: Greece GDP per capita adjusted by PPP from 1990-2014

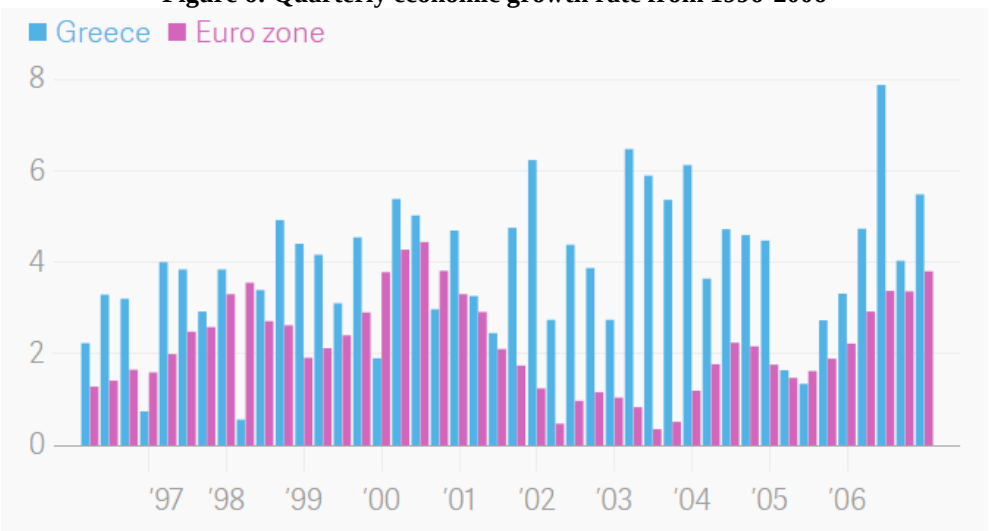


Source: tradingeconomics.com | World Bank

Figure 5: Greece GDP annual growth rate from 1996-2016

Source: tradingeconomics.com | National Statistical Service of Greece

Figure 6: Quarterly economic growth rate from 1996-2006

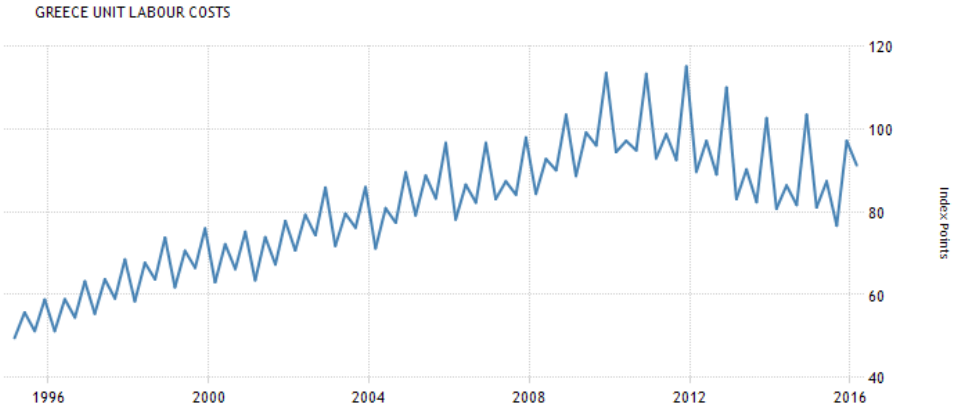


Source: Eurostat

Figure 7: Greece Inflation Rate from 1960-2016

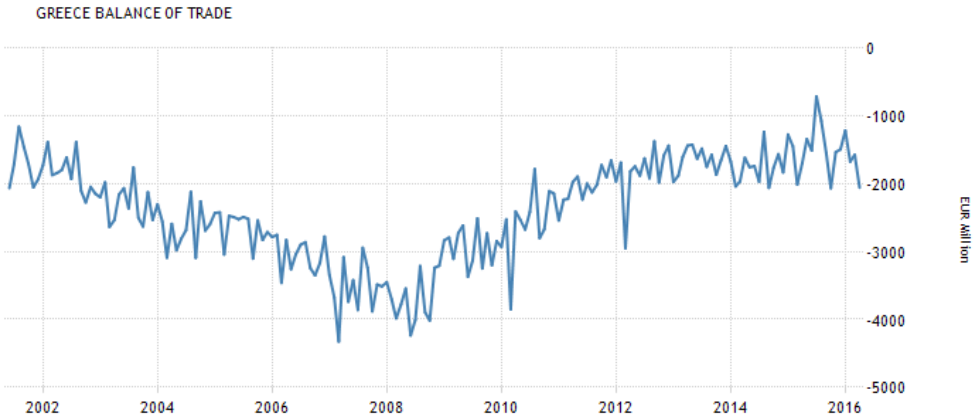
Source: tradingeconomics.com | National Statistical Service of Greece

Figure 8: Greece Unit Labour Costs from 1995 to 2016



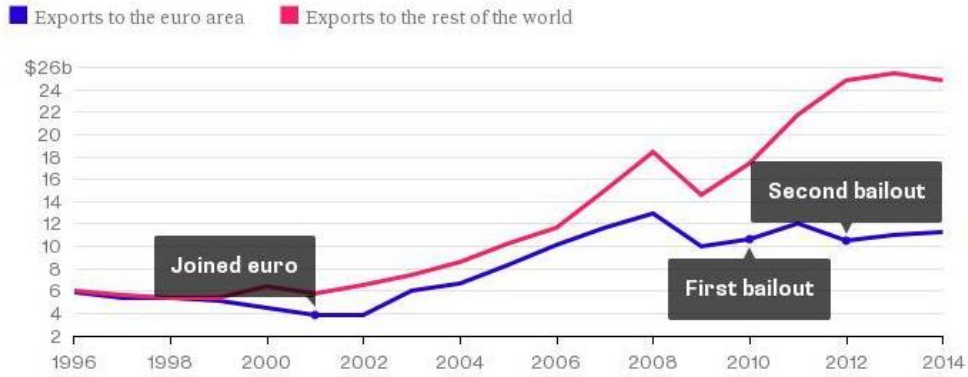
Source: tradingeconomics.com | Eurostat

Figure 9: Greece Balance of Trade from 2001 to 2016



Source: tradingeconomics.com | National Statistical Service of Greece

Figure 10: Comparison of Greek export to EU and rest of the world from 1996 to 2014

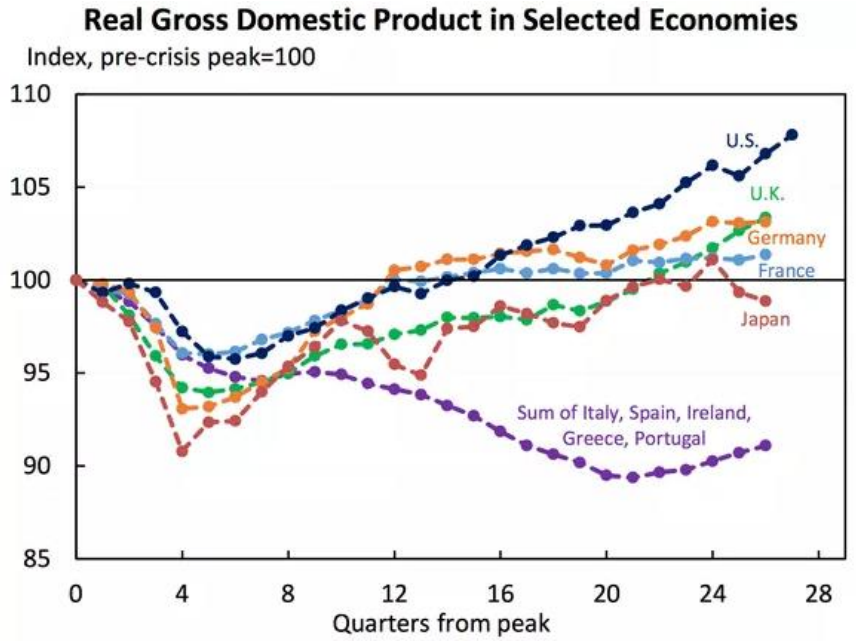


Source: International Monetary Fund

Figure 11: Greece Government Budget as % of GDP from 1995-2013

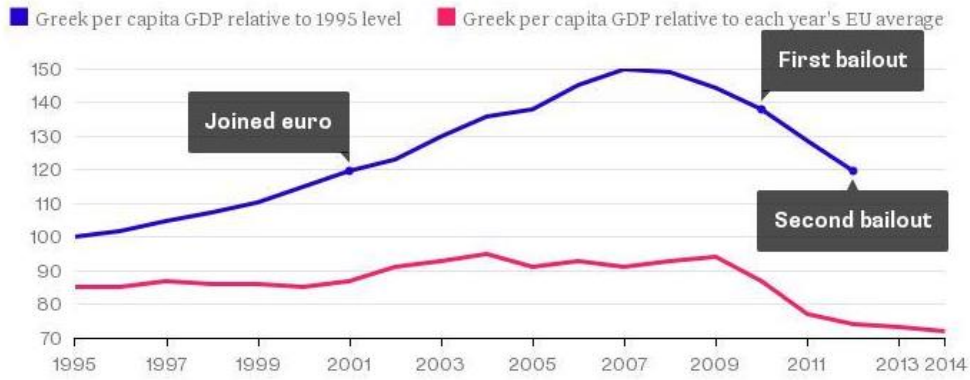
Source: tradingeconomics.com | Eurostat

Figure 12: Change in real GDP in selected economies by quarters since 2007



Source: Council of Economic Advisers

Figure 13: Greece comparative GDP per capita from 1995 to 2014



Source: Eurostat

Figure 14: Greece corporate tax rate from 2006 to 2015

Source: tradingeconomics.com | General Secretariat for Information Systems (GSIS) Greece

Figure 15: Greece unemployment rate from 1998 to 2016

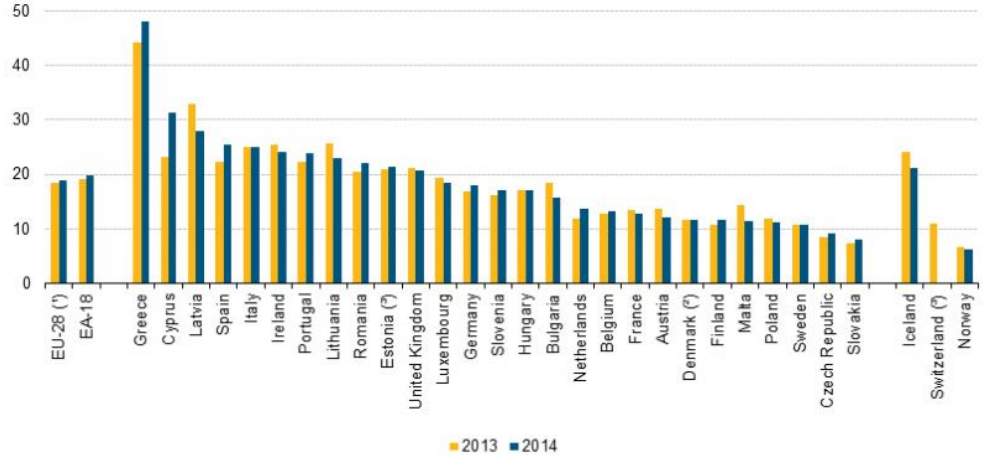
Source: tradingeconomics.com | National Statistical Service of Greece

Figure 16: People at risk of poverty or social exclusion, by age group, 2014

	Total	Children (0–17)	Adults (18–64)	Elderly (65 years and over)
EU-28	24.4	27.8	25.4	17.8
EA-18	23.5	25.6	25.1	16.0
Belgium	21.2	23.2	21.6	17.3
Bulgaria (')	40.1	45.2	36.4	47.8
Czech Republic	14.8	19.5	14.6	10.7
Denmark	17.9	14.5	21.3	10.8
Germany	20.6	19.6	22.0	17.4
Estonia (')	26.0	23.8	24.0	35.0
Ireland	27.4	30.3	29.2	13.0
Greece	36.0	36.7	40.1	23.0
Spain	29.2	35.8	31.8	12.9
France	18.5	21.6	19.9	10.1
Croatia	29.3	29.0	29.3	29.7
Italy	28.3	32.1	30.0	20.2
Cyprus	27.4	24.7	28.3	27.2
Latvia	32.7	35.3	30.0	39.3
Lithuania	27.3	28.9	25.6	31.9
Luxembourg	19.0	26.4	19.4	6.4
Hungary	31.1	41.4	31.5	18.1
Malta	23.8	31.3	21.8	23.3
Netherlands	16.5	17.1	18.9	6.9
Austria	19.2	23.3	18.9	15.7
Poland	24.7	28.2	25.2	18.2
Portugal	27.5	31.4	28.3	21.1
Romania	40.2	51.0	38.7	34.0
Slovenia	20.4	17.7	21.3	20.1
Slovakia	18.4	23.6	18.1	13.4
Finland	17.3	15.6	17.9	17.0
Sweden	16.9	16.7	17.2	16.5
United Kingdom	24.1	31.3	23.2	19.3
Iceland	11.2	13.7	11.0	7.3
Norway	13.5	11.9	15.0	9.9
Switzerland (²)	16.3	17.2	12.7	29.6
FYR of Madeconia	43.2	46.9	43.1	38.4
Serbia	43.1	43.4	44.8	36.6

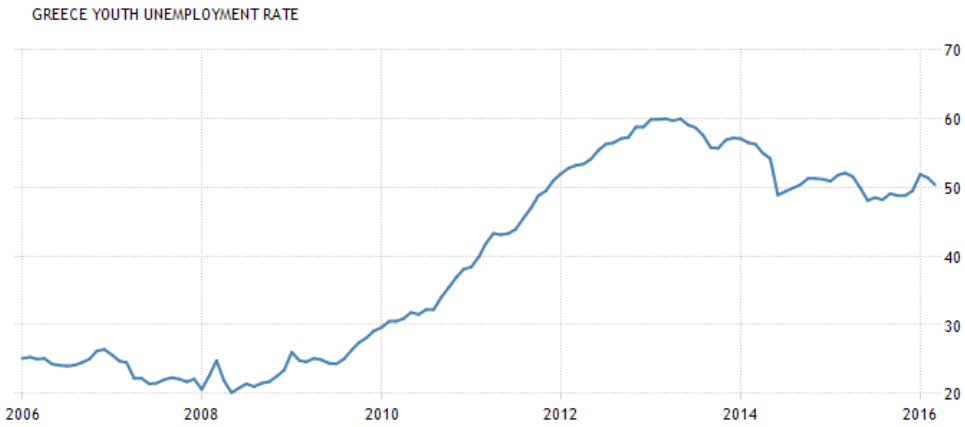
Source: Eurostat

Figure 17: At-risk-of poverty rate anchored at a fixed moment in time (2008)



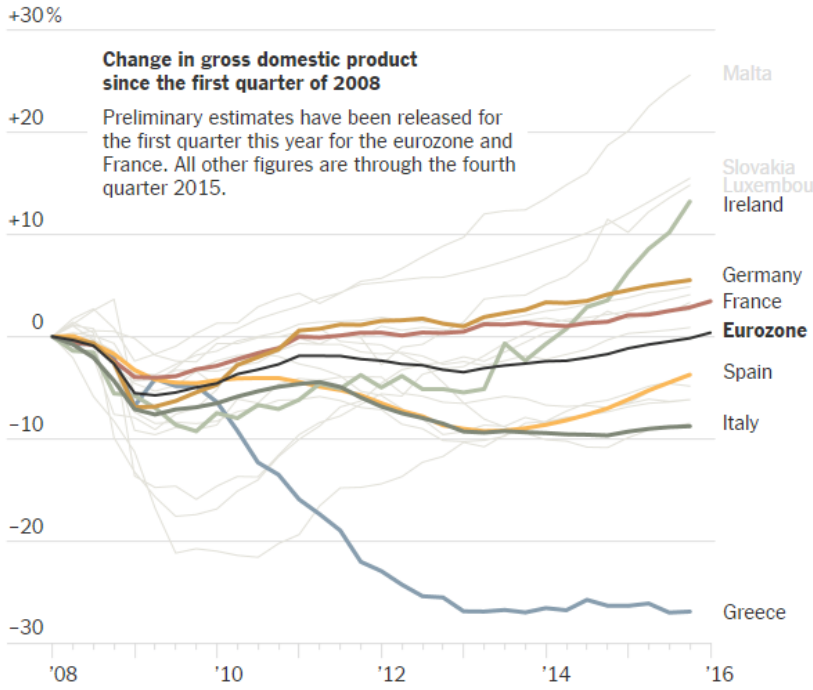
Source: Eurostat

Figure 18: Greece youth unemployment rates from 2006 to 2016



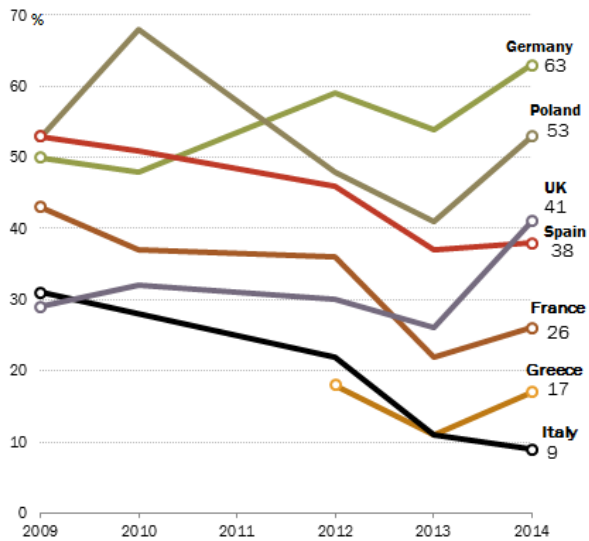
Source: tradingeconomics.com | Eurostat

Figure 19: Europe's change in gross domestic product since first quarter of 2008



Source: New York Times | Eurostat

Figure 20: Percentage saying economic integration of Europe strengthened country's economy



Source: Pew Research Centre

Acknowledgments

I would like to express gratitude to Dr. Visvizi for directing my attention to this explanation of Greece's aspirations in joining the European Communities, the details of which are far more extensively covered in her 2013 publications 'Democracy or Demagogy?' This, among many of her other recommendations, directed me to look beyond the information provided by mainstream literature and at the perspectives of the inquiry ignored by those publications. I am very thankful for her having given me the opportunity to discuss this research inquiry with her, as well as taking the time to read and offer advice regarding my essay. The comments I received in this essay helped me look to better focus the analysis.

I am also very grateful for the guidance I received from Professor Mazzucelli, leading me to the works of Charles A. Kupchan and Kalypso Nicolaidis, which offered thorough explanations for the development of the European narrative and its development into the 21st-century. Her inputs helped structure the content of the essay as well as provide many of the foundations the research was built upon.

Bibliography

Bernanke, Ben S. "Deflation: Making Sure "It" Doesn't Happen Here." National Economists Club. Washington DC. 21 Nov. 2002. *The Federal Reserve Board*. Web. 9 July 2016.

Busch, Klaus. "Is the Euro Failing? Stuctural Problems and Policy Failures Bringing Europe to the Brink." *Friedrich Ebert Stiftung* (2012): n. pag. Friedrich Ebert Foundation, Apr. 2012. Web. 6 July 2016.

"Chapter 2. Crisis of Confidence in the EU Ending?" *Pew Research Centre*. Pew Research Centre, 12 May 2014. Web. 4 July 2016.

Daley, Suzanne. "Greeks Reject Bailout Terms in Rebuff to European Leaders." *The New York Times*. The New York Times, 05 July 2015. Web. 2 July 2016.

Doxiadis, Aristos. "What Greece Needs." *The New York Times*. The New York Times, 26 Feb. 2015. Web. 20 June 2016.

European Union. European Commission. Eurobarometer. *Public Opinion and the euro*. N.p.: n.p., 2015. *Eurobarometer*. Web. 2 July 2016.

European Union. European Commission. Eurostat. *People at Risk of Poverty or Social Exclusion*. N.p.: n.p., 2015. *Eurostat*. Web. 3 July 2016.

European Union. European Commission. Eurostat. *Young people - migration and socioeconomic situation*. N.p.: n.p., 2014. *Eurostat*. Web. 3 July 2016.

European Union. Publications Office. *Introducing the Euro: Convergence Criteria*. N.p.: n.p., n.d. *EUR-Lex*. Web. 21 June 2016.

Ewing, Jack, and Jack Healy. "Cuts to Debt Rating Stir Anxiety in Europe."*New York Times*. N.p., 27 Apr. 2010. Web. 20 June 2016.

Fabbrini, Federico. "The Fiscal Compact, the "Golden Rule," and the Paradox of European Federalism." *Boston College International and Comparative Law Review* 36.1 (2013): n. pag. Boston College Law School, 26 Mar. 2013. Web. 7 July 2016.

Goodman, Peter S. "Europe's Economy, After 8-Year Detour, Is Fitfully Back on Track." *The New York Times*. The New York Times, 29 Apr. 2016. Web. 22 June 2016.

"Greece Balance of Trade | 2001-2016 | Data | Chart | Calendar | Forecast."*Trading Economics*. National Statistical Service of Greece, n.d. Web. 23 June 2016.

"Greece Corporate tax Rate | 1981-2016 | Data | Chart | Calendar | Forecast."*Trading Economics*. GSIS Greece, n.d. Web. 23 June 2016.

"Greece Government Budget | 1995-2016 | Data | Chart | Calendar | Forecast."*Trading Economics*. Eurostat, n.d. Web. 25 June 2016.

"Greece Inflation Rate | 1960-2016 | Data | Chart | Calendar | Forecast."*Trading Economics*. National Statistical Service of Greece, n.d. Web. 24 June 2016.

"Greece - OECD Data." *OECD Data*. OECD, 2016. Web. 2 July 2016.

"Greece Unemployment Rate | 1998-2016 | Data | Chart | Calendar | Forecast."*Trading Economics*. National Statistical Service of Greece, n.d. Web. 24 June 2016.

"Greece Unit Labour Costs | 1995-2016 | Data | Chart | Calendar | Forecast."*Trading Economics*. Eurostat, n.d. Web. 23 June 2016.

"Greece Youth Unemployment Rate| 1998-2016 | Data | Chart | Calendar | Forecast."*Trading Economics*. Eurostat, n.d. Web. 24 June 2016.

Heise, Michael. "The Real Tragedy of the Greek Debate." *World Economic Forum*. N.p., 25 June 2015. Web. 1 July 2016.

"Interest Rates - Long-term Interest Rates - OECD Data." *OECD Data*. OECD, 2016. Web. 25 June 2016.

Kirkegaard, Jacob Funk. "Economic Crisis: The Global Impact of a Greek Default." *Peterson Institute for International Economics* (2015): n. pag. 25 June 2015. Web. 30 June 2016.

Lesser, Ian O., Stephen Larrabee, Michelle Zanini, and Katia Vlachos-Dengler. "Chapter Two: Greece's New Strategic Environment." *Greece's New Geopolitics*. Santa Monica, CA: Rand, 2001. 7-37. Print.

Magoulios, George, and Stergios Athianos. "The Trade Balance of Greece in the Euro Area." *South-Eastern Europe Journal of Economics* 21 (2013): n. pag. 2013. Web. 24 June 2016.

Moravcsik, Andrew. "The European Constitutional Compromise and the Neofunctionalist Legacy." *The European Constitutional Compromise and the Neofunctionalist Legacy* 12.2 (2005): 349-86. *Journal of European Public Policy*. Princeton, 2 Apr. 2005. Web. 20 June 2016.

OECD. Center for Tax Policy and Administration. *Revenue Statistics 2014- Greece*. OECD, 2014. *OECD*. Web. 28 June 2016.

Paris, Costas. "Greek Shipping Industry Extends Its Dominance." *WSJ*. N.p., 9 Aug. 2015. Web. 4 July 2016.

Rasku, Minna. *On the Border of East and West: Greek Geopolitical Narratives*. Ed. Jussi Kotkavirta. Jyväskylä: U of Jyväskylä, 2007. Web 25 June 2016,

Sakellariou, Alexandros. "Golden Dawn and Its Appeal to Greek Youth."*Golden Dawn and Its Appeal to Greek Youth* (2014): n. pag. July 2014. Web. 4 July 2016.

Sarotte, Mary E. "Eurozone Crisis as Historical Legacy: The Enduring Impact of German Unification, Twenty Years On." *Crisis in the Eurozone Transatlantic Perspective*. New York: Council on Foreign Relations, International Institutions and Global Governance Program, 2010. N. pag. Print.

Sfakianakis, John. "The Cost of Protecting Greece's Public Sector." *The New York Times*. The New York Times, 10 Oct. 2012. Web. 27 June 2016.

Shambough, Jay C. "The Euro's Three Crises." *Brookings Papers on Economic Activity* (2012): n. pag. Georgetown University, 2012. Web. 6 July 2016.

Stokes, Bruce, and Sara K. Goo. "5 Facts about Greece and the EU." *Pew Research Centre*. N.p., 27 Jan. 2015. Web. 27 June 2016.

Tartar, Andre. "Here's What Membership in the Euro Did for Greece."*BloombergMarkets*. Bloomberg, 17 July 2015. Web. 21 June 2016.

"The Eurozone Crisis and the Rise of Soft Euroscepticism in Greece." *Ballots Bullets School of Politics International Relations University of Nottingham*. The University of Nottingham, 26 Feb. 2013. Web. 20 June 2016.

"The New Sick Man of Europe: the European Union" *Pew Research Centre*. Pew Research Centre, 13 May 2013. Web. 4 July 2016.

Traynor, Ian. "Angela Merkel Dashes Greek Hopes of Rescue Bid." *The Guardian*. Guardian News and Media, 11 Feb. 2010. Web. 20 June 2016.

Troiti ñ, David R. 'De Gaulle and the European Communities', *Institute for European Studies* (2008): 139-52. Web. 22 June 2016.

Nelson, Belkin, and Mix. United States. U.S. Congressional Research Service. *Greece's Debt Crisis: Overview, Policy Responses, and Implications*. N.p.: n.p., 2011. *Congressional Research Digital Collection*. Web. 22 June 2016.

Visvizi, A. (2012) 'The June 17 Elections in Greece: Domestic and European Implications', PISM Policy-Paper, No. 31, June 2012, Polish Institute of International Affairs (PISM), Warsaw, Poland.

Visvizi, A. (2013) Democracy or demagogy? The Greek political actors on the sovereign debt crisis, in: Rye, L. (ed.) Distant voices. Ideas on democracy and the Eurozone crisis, Rostra Books – Trondheim Studies in History, Trondheim, Norway: Akademika Publishing, pp. 23-50.

Wodak, Ruth, Majid Khosravinik,and Brigitte Mral. *Right-Wing Populism in Europe*. N.p.: Bloomsbury Academic, 2013. Web. 12 July 2016.

East European Jewish Children's Health Conditions on the Lower East Side, New York City, 1890-1914

Yibing Du

Author background: Yibing Du grew up in China and currently attends The High School Affiliated to Renmin University of China, located in Beijing, China. Her Pioneer seminar topic was in the field of history and titled "Topics in the History of Public Health."

Introduction

From 1890 to 1914, large numbers of immigrants entered America, including 2,021,000 East European Jews.[63] As these Jews entered the United States, physicians and officers evaluated their physical health, and those identified as healthy received permission to enter the country. Similar to other immigrant groups, many of these Jewish immigrants possessed no professional skills; distinct from other groups, they had a large proportion of women and children and thus a high percentage of permanent immigrants.

Upon arrival, East European Jews joined their co-religionists who had lived in America for a few decades, most originally from Western Europe. In fact, from 1880 to 1930, 230,000 Jews in the United States absorbed 3,000,000 Jewish immigrants.[64] For years, the earlier immigrants, a considerable proportion of whom were from Germany, "labored on to win acceptance and integration within the larger society,"[65] and as a result, many of them had already become economically self-sufficient and culturally Americanized by 1880.

Many of the East European Jewish immigrants who entered the U.S. through the port of New York City settled on the Lower East Side, close to their arrival spot on Ellis Island. There, many lived in crowded buildings and worked in dusty factories or workshops. Even so, surprisingly, Jewish immigrants (especially children) tended to have a lower mortality rate than other immigrants or even natives; this pattern also existed among Jews in Europe.[66]

Although there were general hospitals in big cities, a sizeable percentage of assimilated Jews became dissatisfied with the medical care provided in these institutions. Many general hospitals did not serve kosher food, some lacked staff who spoke Yiddish, and some, from time to time, attempted to convert Jews. Thus, in larger cities, Americanized Jews often created their own medical associations, including hospitals. In the case of New York City, the first Jewish hospital, Mount Sinai Hospital, was founded as early as 1853.[67] Three decades later, significant changes in the demands upon the charitable organizations and

[63] Deborah Dwork, "Health Conditions of Immigrant Jews on the Lower East Side of New York: 1880-1914," *Medical History* 25, no.1 (1981): 1-2, doi: 10.1017/S0025727300034086.
[64] Graenum Berger, "American Jewish Communal Service 1776-1976: From Traditional Self-Help to Increasing Dependence on Government Support," *Jewish Social Studies* 38, no.3/4 (1976): 233, http://www.jstor.org/stable/4466936.
[65] Naomi W. Cohen, *Encounter with Emancipation: The German Jews in the United States 1830–1914* (Philadelphia: The Jewish Publication Society of America, 1984), xi, EBook, Varda Graphics, Inc.
[66] Dwork, "Health Conditions of Immigrant Jews": 28.
[67] Cyrus Adler ed., *The American Jewish Year Book* (Philadelphia: The Jewish Publication Society of America, 1899), 212, http://www.ajcarchives.org/AJC_DATA/Files/1899_1900_5_LocalOrgs.pdf.

medical facilities in the Jewish community (mainly among German Jews) took place with the arrival of new immigrants, due to the fact that this larger East European population, facing bad living and working conditions, often needed medical care.

At first, many German Jews did not welcome new immigrants, since they feared that a large Jewish immigrant population would provoke anti-Semitism.[68] Nevertheless, once they realized that they could not stop new immigrants from flocking in, they set up numerous philanthropic organizations, including the Russian Emigrant Relief Fund in 1881 and the Hebrew Emigrant Aid Society in 1881.[69] These organizations offered daily supplies, basic medical treatments, and midwifery services. At the same time, Jewish medical associations, including the Mount Sinai Hospital, expanded and accepted patients who were too poor to afford to pay for care out of a sense of duty.

Jewish women, especially those interested in children and health, accomplished many philanthropic works. Many middle class Jewish women, who had arrived before the wave of the 1880s, were particularly active in philanthropic work that sought to ameliorate the health crisis of the Jewish community. They created associations that sponsored philanthropic, religious, and community events to raise funds for hospitals and children's asylums and raised money to support those associations – which were philanthropic and social.[70] Many Americanized Jewish women who received nursing education dedicated themselves to teaching new immigrant women basic nursing skills and to organizing nursing services, and several groups of social workers accelerated this progress. For example, Lillian Wald established the Henry Street Settlement House that offered nursing's training services.[71]

Besides women-oriented Jewish philanthropic groups, evidence indicates that there were several factors that account for the surprisingly low children's mortality rate in the Jewish community. This paper will argue that the philanthropic activities of assimilated Jews made a very large difference in the health of East European Jewish children's health in the crowded New York City neighborhoods during the era of massive Jewish immigration. In particular, this paper will demonstrate that the various medical institutions operated by established Jews contributed to lowering Jewish children's mortality rates. This paper will emphasize how Americanized Jews helped out new immigrants from Eastern Europe, how the latter reacted to their benefactors, and how this complicated relationship influenced medical and charitable organizations.

Health Problems East European Jews Faced: A Critical Issue

Some 675,000 East European Jews entered the United States between 1881 and 1900 and that number doubled to approximately 1,346,000 between 1901 and 1914.[72] During this era, Jews living in Eastern European countries (including Russia and Poland) were not only suffering from poverty but also being persecuted by the Czar. While some Jews lived in significant sized cities, regulations forced many of them to live in poor towns (*shtetlach*) in the Pale of Settlement, located in western Russia and a part of Poland that was controlled by

[68] Irving A. Mandel, "Attitude of the American Jewish Community Toward East-European Immigration: As Reflected in the Anglo-Jewish Press (1880-1890)," *American Jewish Archives* (1950): 13, http://americanjewisharchives.org/publications/journal/PDF/1950_03_01_00_mandel.pdf.
[69] Mandel, 28; Berger, "American Jewish Communal Service 1776-1976," 233.
[70] Beth S. Wenger, "Jewish Women of the Club: The Changing Public Role of Atlanta's Jewish Women (1870-1930)," *American Jewish History* 76, no. 3 (1987): 312, 315, http://www.jstor.org/stable/23883223.
[71] Marjorie N. Feld, *Lillian Wald: A Biography* (Chapel Hill: The University of North Carolina Press, 2008), 2, Kindle Edition.
[72] Dwork, "Health Conditions of Immigrant Jews": 1-2.

the Czar.[73] Under such pressure, East European Jews sought to emigrate in order to escape the miserable political and economic environment of the Czar's rule and the government's persecution. As one commentator noted in 1905, while "the anti-Semite spirit prevents them from gathering together in any great numbers in Europe," America was relatively tolerant of immigrants.[74]

As the East European Jews immigrated to the U.S., they were distinct from other groups because of their large population, tendency to settle in urban areas, permanency of migration, high percentage of women, rapidly declining fertility, and strong family cohesion.[75] In the eyes of Americans, East European Jews were generally less educated and more religious than their counterpart in other parts of Europe. Communicating in Yiddish and barely speaking English, they had a wide cultural gap from other Americans, which could lead to a lot of discrimination. As new immigrants including East European Jews entered the United State in large numbers, Americans gradually formed prejudice-driven distinctions between the newcomers and the old immigrants, Jews from German or Northern Europe as well as Christians. These distinctions resulted in negative perception of new immigrants. Many Americans claimed that new immigrants were "destitute and uneducated, […] less assimilable and far more troublesome than their 'old' counterparts."

The demographics of these East European Jewish immigrants showed a distinct pattern "[i]n the low proportion of temporary migrants, in the large proportion of women and children, and of immigrants without occupation, as well as in the unequaled proportion of skilled laborers"[76] when compared to their gentile counterparts. Of the 1.2 million Jewish immigrants arriving in the United States between 1899 and 1914, about a quarter were children up to the age of 14 and forty-five percent were women.[77] This fact correlates with the permanence of the Jewish immigration, since scholars pointed out that "[t]he larger the number of women and children among migrants, the more permanent is that migration."[78]

Upon their arrival, all immigrants received a series of health inspections at Ellis Island and other American immigration reception centers. Of those who passed or recovered after brief treatments, many settled in New York City. The Lower East Side area offered these new tenants poor living conditions. In 1915, the Lower East Side of Manhattan had the highest density of population per acre – 696.7 – in contrast to the average density in New York City generally, which was 57.8.[79] Because of the ratio of population to housing, most immigrants lived in congested places. As described in an 1894 report, "[o]vercrowded tenement houses, deficient sanitary conveniences, inadequate street cleaning and house inspection" were common problems in Lower East Side neighborhoods.[80]

---

[73] Stephen H. Norwood and Eunice G. Pollack, ed., "East European Jewish Immigration," *Encyclopedia of American Jewish History* (Santa Barbara: ABC-CLIO, Inc., 2008), 42, EBook, World Wide Web.

[74] Felix Klein, *In the Land of the Strenuous Life* (Chicago: A. C. McClurg& Co., 1905), 37, Google Book.

[75] Jacob J. Lindenthal, "'Abi Gezunt': Health and the Eastern European Jewish Immigrant," *American Jewish History* 70 no. 4 (1981): 437, http://www.jstor.org/stable/23881910.

[76] Liebman Hersch, "International Migration of the Jews," *International Migrations*. New York, 2 (1931): 506, http://www.nber.org/chapters/c5117.

[77] Hersch, 485.

[78] Hersch, 516.

[79] The City of New York Department of Street Cleaning, *Annual Report* (1915), 12, http://archive.org/cletails/annualreportofde00newy_5.

[80] University Settlement Society, *Report of the Year's Work* (New York: Concord Printing Co., 1894), 8, http://archive.org/details/reportuniversityOOuniv.

For those Jewish children who did not have a proper home, they had no choice but to live on the streets. Jacob A. Riis, a famous journalist who conducted investigations on the Lower East Side, recorded his observation: "[e]ven the alley is crowded out. Through dark hallways and filthy cellars, crowded, as is every foot of the street, with dirty children, the settlements in the rear are reached."[81] Children who had a home and a family to live with, who were the vast majority, dwelled "in dark, dreary apartment buildings called tenements. Large families were crammed into one or two rooms, with little heat in winter and no windows to provide a breeze in the heat of summer."[82] The New York City Board of Health reported that such crowded housing could pose great threats to children, because if one of their parents had tuberculosis, the children could be infected easily: "[w]here the parents are affected with tuberculosis the children from the earliest moments of life are exposed to the disease under the most favorable conditions for its transmission."[83] What was worse, the labor organization, the Workmen's Circle, suggested that after infection, children could be constantly afflicted by tuberculosis, reasoning that "[t]uberculosis is a chronic disease. The seeds sown in childhood may not bear fruit […] until adult life. The disease may be latent (asleep) for many years before it 'wakes up' and becomes discoverable."[84]

The sanitation in such housing conditions was poor indeed, and thus epidemics could spread easily. Analyzing the statistics of home locations and contagious diseases on the Lower East Side around 1900, scholars found that there was a notable "relationship between housing and infectious diseases, especially tuberculosis and diphtheria."[85] Dwelling in these places, Jews were exposed to a wide range of potential health risks. For children, such risks could easily lead to illnesses. For example, in Bertha Fox's childhood memory as a girl living on the Lower East Side, she recalled that she and her brother suffered from many illnesses, partly due to their housing conditions: "[t]he interior side of the wall facing the street was damp and often dirty, so the children would indeed get sick. My brother and I would go from one illness to the next, scarlet fever, diphtheria, not to mention all the other childhood illnesses."[86]

Because of the prevalence of anti-Semitism, and probably anti-immigrant attitudes more generally, many Americans did not care much about the sanitation issues on the Lower East Side. Regarding the East European Jews as an inferior group, they were reluctant to offer help. Nativists had negative attitudes towards immigrants and blamed lots of the immigrant groups for bringing certain contagious diseases and worsening social conditions in the city. For example, Jews were deemed as vulnerable to trachoma, tuberculosis, and

---

[81] Jacob A. Riis, *How the Other Half Lives: Studies Among the Tenements of New York* (New York: Trow's Printing and Bookbinding Company, 1890; Project Gutenberg, 2014), 106, http://www.gutenberg.org/ebooks/45502.

[82] Richard Worth, *Immigration to the United States: Jewish Immigrants* (New York: Fact on Files, Inc., 2005), 41, Ebook, World Wide Web.

[83] Board of Health of the City of New York, *The Action of the Health Department in Relation to Pulmonary Tuberculosis and the Scope and Purpose of the Measures Recently Adopted for its Prevention* (Albany and New York: Wynkoop Hallenbeck Crawford Co., 1897), 13, Google Book.

[84] *Arbeiter ring sanitarium, liberti, niuiork, onveizungentsupatsientn*, Bund Archives, Workmen's Circle Papers, Box 10, Folder 57, 3, as cited by Daniel Bender, "'A Hero... for the Weak': Work, Consumption, and the Enfeebled Jewish Worker, 1881-1924," *International Labor and Working-Class History* 56 (1999): 8, http://www.jstor.org/stable/27672593.

[85] Dwork, "Health Conditions of Immigrant Jews": 23.

[86] Bertha Fox, "The Movies Pale in Comparison," in *My Future Is in America: Autobiographies of Eastern European Jewish Immigrants*, ed. Jocelyn Cohen et al. (New York and London: New York University Press): 210, ProQuest ebrary.

neurasthenia, which suggested that they were inferior to Americans.[87] The ideology of anti-immigration stigmatized immigrants "as the etiology of a wide variety of physical and societal ills," and "anti-immigrant rhetoric and policy have often been framed by an explicitly medical language."[88] Markel contended, "If authorities and anti-immigration advocates found that one classification failed to reject the 'most objectionable', they soon created a new one that emphasized contagion, mental disorder, chronic disability, or even a questionable physique."[89] Some Americans proudly defended the anti-immigrant attitude by justifying this as a process of selection, implying that turning down the sick and the weak was reasonable. "It is a mistake to think that this country is being made a dumping-ground for Europe's rubbish. Year by year we are acquiring, by a process of natural selection, the pick of the nations."[90]

Governmental departments in charge of health issues had no intention of solving the sanitation problem in the Lower East Side, partly because of their anti-Jewish inclination. A letter of complaint published in the *New York Times* expressed concern and indignation: "[t]he condition of the streets on the Lower East Side is beyond description. Never in its history has a Street Cleaning Superintendent displayed such wanton disregard for its welfare, and it is only a question of a little time when the health of its residents will be seriously endangered."[91] In one case in 1907, Mayor McClellan, after receiving several requests to tackle the sanitation problem on the Lower East Side, asked the Street Cleaning Department to deal with this situation. However, the Department's action was so perfunctory that the *New York Times* pointed out that "[t]here was much scurrying around, but nothing in the way of relief was accomplished."[92] Such neglect was partly based on prejudice and fear. Government officials tended to have stereotypical views about certain areas of the city: "[t]he health officers call the Tenth the typhus ward; in the office where deaths are registered it passes as the 'suicide ward,' for reasons not hard to understand."[93]

Meanwhile, few immigrants trusted the Health Department and governmental organizations in general, so most chose not to report their diseases because they thought they would be expelled. As Riis noted, typhus fever and small-pox "sprout naturally among the hordes that bring the germs with them from across the sea, and whose first instinct is to hide their sick lest the authorities carry them off to the hospital to be slaughtered, as they firmly believe."[94]

Low East European Jewish Children Mortality Rates: A Surprising Pattern

In lots of cases, immigrants entering the United States around 1900 struggled with diseases and overall sanitary conditions. However, the trend among Jewish immigrants

[87] H. Markel and A. M. Stern, "Which Face? Whose Nation? Immigration, Public Health, and the Construction of Disease at America's Ports and 1891-1928," *American Behavioral Scientist* 42 No. 9 (1999): 1328, doi: 10.1177/0002764299042009010.
[88] Markel, “The Foreignness of Germs”: 757.
[89] Markel, 758.
[90] Arthur Henry, "Among the Immigrants," *Scribner's Magazine*, January – June, 1901, 29, 311. https://babel.hathitrust.org/cgi/pt?id=mdp.39015030597127;view=1up;seq=5.
[91] "Some Neglected Streets: Complaint of the Condition of Thoroughfares on the Lower East Side," *New York Times*, September 1, 1901, 8, ProQuest Historical Newspapers.
[92] "Mayor Acts to Stop Peril from Refuse," *New York Times*, June 29, 1907, 1, ProQuest Historical Newspapers.
[93] Riis, *How the Other Half Lives*, 110.
[94] Jacob A. Riis, *The Children of the Poor* (New York: Charles Scribner's Sons, 1892; Gutenberg Project, 2010), 109, http://www.gutenberg.org/ebooks/32609.

worldwide turned out to be rather different. The conspicuous contrast between the terrible housing conditions and the low death rates in Jewish neighborhoods began to draw outside attention starting in the 1890s. As Riis recorded his observation about Jewish tenements:

> [T]he tenement seems to be unusually bad even for that bad spot; but when we came to look up its record, from the standpoint of the vital statistics, we discovered that not only had there not been a single death in the house during the whole year […] I had met with similar experiences, if not quite so striking, often enough to convince me that poverty and want beget their own power to resist the evil influences of their worst surroundings.[95]

The surprisingly low mortality rates were particularly significant when it came to Jewish children. According to the *1890 U.S. Census Report for New York City and Brooklyn*, the mortality rates of those whose mothers were born in Russia and Poland (most of whom were Jewish) were the lowest of all recorded nationalities (1,485 per 100,000). For Jewish immigrants in general, child mortality was also as low as 2,867 per 100,000.[96]Jews living in European cities also showed unusually low child death rates, including Budapest in 1885-1893, Prussia 1888-1897, Tunis 1894-1900, Berlin 1905, Vienna 1901-1903, London 1901-1906, and from 1896-1900 in Cracow.[97]

Scholars often regarded the good sanitary habit, rare alcoholism, and strong philanthropic tradition in Jewish communities as the reasons for the unusual phenomenon in the Jews' low mortality rate.[98] However, good sanitary habit was least correlated to children's mortality rates. Contrary to popular belief, Jewish sanitary habits did not necessarily make them healthier. Although some people once considered eating kosher food as a means to prevent tuberculosis from being ingested, this theory proved wrong as people realized tuberculosis spread via inhalation.[99] Some literature extolled Jewish households for "their unusual cleanliness,"[100] and Jewish housewives were regarded as expert at cleaning rooms and kitchen,[101] which may have benefited children's health. However, in general, even though the poor might be able to improve their households bit by bit, unless they became economically better off, they could not avoid bad sanitation completely. Good sanitation was not really possible for the poor, thus it didn't account for low mortality rates among Jewish residents.

The second factor that affected children's death rate was low alcoholism. Alcohol consumption could take up a large proportion of the family's financial resources and thus limit the family's health care budget. The 1915 *Report on the Cost of Living for an Unskilled Laborer's Family in New York City* estimated that a family of five needed 20 dollars (2.4%) a year to cover routine health care.[102] For the underprivileged, inflated prices in New York

---

[95] Ibid., Riis, 43.
[96] *U.S. Census Report for New York City and Brooklyn*, 1890, as cited by Dwork, 27.
[97] Dwork, 28.
[98] G. A. Condran and E. A. Kramarow, "Child Mortality among Jewish Immigrants to the United States," *The Journal of Interdisciplinary History* 22 No. 2 (1991): 230, http://www.jstor.org/stable/205867.
[99] Condran, 229.
[100] Condran, 231.
[101] Maurice Fishberg, "Health and Sanitation of the Immigrant Jewish Population of New York," *The Menorah* 33, no. 2 (1902): 73-74.
[102] *Report on the Cost of Living for an Unskilled Laborer's Family in New York City*, 1915. As cited by Alan M. Kraut, "Healers and Strangers: Immigrant Attitudes toward the Physician in America – A Relationship in Historical Perspective," *JAMA* 263, no. 13 (1990): 1809, doi: 10.1001/jama.1990.03440130095032.

City, as the *Los Angeles Times* reported, were so drastic that it made eggs "too high-living even for sick people."[103] Therefore, rare alcoholism influenced children's death rates because limited alcohol consumption saved money for families to apply to daily nutrition and medical care in case the children fell ill. Moreover, parents with drinking issues were less likely to get well-paid jobs. Nor were they capable of taking good care of their children.

Among the unique features of the Jewish community, the one that directly lowered the children's chance of dying was their philanthropic tradition. Americanized Jews lent a helping hand to new immigrants by investing large amounts of money in "orphanages, day nurseries, training schools, free medical clinics, hospitals, and old age homes"[104] to help newly arrived Jews contact their family members, find jobs, and obtain basic welfare care.

American Jews and Their Philanthropic Medical Associations: A Helping Hand

The interactions between East European Jewish immigrants and their fellow acculturated Jews played a vital role in assisting many fellow Jews with nutrition, improving living conditions, and providing medical care for children. With limited professional skills and facing language barriers, most East European Jews were not qualified for middle- or high-income jobs or support their families, and had trouble even finding low-paying jobs. Although many received help from relatives who came a short time ahead of them, they were still in need of medical facilities and the nursing that American Jewish charities rendered. Therefore, they were very much dependent on the support from their already Americanized fellow Jews, especially when they first arrived. In 1912, a guideline for new Jewish immigrants listed trustworthy charitable societies that included the Hebrew Sheltering and Immigrant Aid Society, the Clara de Hirsch Home for Immigrant Girls, the Industrial Removal Office, and the Young Women's Hebrew Association.[105] In many institutions like these, support for the sick poor mainly took the form of philanthropic work, initiated by better-off Jews. These volunteers made extensive efforts to address the hygiene issues facing the Lower East Side since the government did a poor job. For instance, social workers in a Jewish neighborhood association, the University Settlement Society, claimed that they had worked to secure "the more efficient service of the Board of Health and Street Cleaning Departments, [before government could] provide for the people all those legitimate contributions to health and right living."[106]

Among American Jewish social workers, health issues among Jewish children generated a great deal of concern, given that the miserable living conditions were an underlying cause of illnesses or even epidemics. Thus, American Jewish social workers founded new hospitals, expanded Children's Departments in existing hospitals, offered home-visit nursing, and opened up maternity homes in order to provide medical care for sick Jewish children.

The opening of new Jewish hospitals made good medical resources more reachable for new immigrants, and thus improved their health. Middle-class women, with the help of young physicians, created new medical institutions. These community-minded women set up philanthropic women's organizations primarily to raise funds for the sick and the poor.

---

[103] "The Cost of Living," *Los Angeles Times*, May 13, 1917, 12, ProQuest Historical Newspapers.
[104] Norwood ed., "German Jews in America," *Encyclopedia of American Jewish History*, 35.
[105] John Foster Carr, *Guide to the United States for the Jewish Immigrants: A Nearly Literal Translation of the Second Yiddish Edition* (New York: The Connecticut Daughters of the American Revolution, 1912), 9-10, 20, Google Book.
[106] University Settlement Society, *Report of the Year's Work* (New York: Concord Printing Co., 1894), 8, http://archive.org/details/reportuniversityOOuniv.

At the same time, anti-Semitism directed at Jewish physicians deprived them of the right to practice medicine in many general hospitals. When rejected by general hospitals, young Jewish physicians looked for a place to use their skills and tried in various ways to do so. Scholars, A.M. Kraut and D. A. Kraut, noted that:

> Because Jewish physicians were often denied staff privileges at voluntary charitable hospitals operating under religious auspices, they began to open dispensaries, the precursors of outpatient clinics, in order to practice medicine. These dispensaries often expanded to include a small annex with beds; thus, many became clinics for the poor and small proprietary hospitals for one or several physicians.[107]

These newly founded Jewish hospitals were not only medically creditable but also culturally friendly, and hence, new immigrants who resisted treatment in general hospitals could receive good medical care there.

At the end of the nineteenth century, many smaller medical institutes were being set up. In the meanwhile, two large Jewish hospitals, the Mount Sinai Hospital and the Beth Israel Hospital in New York City, accepted patients regardless of ethnic and religious background and acted as philanthropic institutions for the poor. Founded in 1853, the Mount Sinai Hospital of the City of New York, defined its own goal as "[t]o treat the sick (free of charge to the poor) without distinction as to creed, color or nationality."[108] The relatively new Beth Israel Hospital was established on December 1, 1889. Its objective, according to *The American Jewish Year Book,*[109] was "[t]he maintenance of a hospital and dispensary in the Down Town East-Side district of New York City for the purpose of affording medical and surgical relief to the sick poor of the Jewish faith residing in said district, but not excluding other sects from its benefits."[110] Thus, the hospitals served both the Jews and the Gentiles, the affluent and the needy. Because Jewish children from poor families gradually came to have the same right to receive treatment as their better-off counterparts, their risks of being afflicted by illnesses were reduced.

The Children's Department of Mount Sinai Hospital had expanded in size and diversified in service significantly since its opening in 1899. It hired more physicians and assistants, offered more services, and provided larger numbers of patients with consultations and prescriptions. In 23 years, the size of the staff in the Children's Department increased fourfold: from 11 (including 3 Physicians and 8 Assistants) in 1890 to 44 (including 9 Chiefs, 8 First Assistants, and 27 Clinical Assistants) in 1913.[111]

Meanwhile, the Mount Sinai Hospital's Children's Department gradually broadened the range of services it provided, as recorded in its Annual Reports. In 1910, to provide follow-up medical care, the Children's Department started "visiting the homes of children and infants discharged from the Hospital, for the purpose of instructing their mothers in the

---

[107] A. M. Kraut and D. A. Kraut, *Covenant of Care: Newark Beth Israel and the Jewish Hospital in America* (New Brunswick: Rutgers University Press, 2006), 23-24, ProQuest ebrary.
[108] Adler ed., *The American Jewish Year Book*, 194.
[109] Ibid., Adler ed., 194.
[110] Ibid., Adler ed., 212.
[111] Mount Sinai Hospital, *Annual Report of the Mount Sinai Hospital of the City of New York* (New York: Press of Stettiner Brothers, 1900), 33, http://archive.org/details/annualreport1899moun. Mount Sinai Hospital, *Sixtieth Annual Report of the Mount Sinai Hospital of the City of New York* (New York: Press of Stettiner Brothers, 1913), List of Dispensary Staff, https://archive.org/details/annualreport1910moun.

principles of hygiene, methods of infant feeding and sanitation."[112] In 1911, the hospital planned to set up independent Children's Clinics "for a reduction of infant mortality, directly traceable to this undertaking." [113] Later in 1912, the Department "employed a Kindergaertnerin, who spent several hours every afternoon with those children who were not confined to their beds."[114] Aware of the potential for contagious infection, in 1913, the hospital used its dispensary building to "meet an urgent need in the Children's Department by forming a clinic for the treatment of vaginitis, an infectious condition which is so common among the children of the poor as to represent a public menace."[115] Moreover, to more thoroughly help the young patients and their families, the Children's Department cooperated with the Hebrew Orphan Asylum, the Hebrew Infant Asylum, the United Hebrew Charities, an the Sanitarium for Hebrew Children to help poor patients connect to these charitable societies.[116] The institution helped poor patients' families, so that children could still be fed when their parents fell ill: "[w]hen the breadwinner of the family is detained in the ward, it is of immeasurable help to the physician if the patient knows that his family is not starving at home."[117]

From 1899 to 1909, as illustrated in Figure 1, the numbers of children consultations and prescriptions Mount Sinai Hospital respectively increased from 9409 to 25,000, and 9435 to 25,469. (In 1909, the hospital only collected data of 7 months, thus the number of children treated over the whole year is estimated to be 25,000 and 25,469, respectively.) As a result, an increasingly large number of Jewish children benefitted from the Children's Department of Mount Sinai Hospital.

[112]Mount Sinai Hospital, *Fifty-seventh Annual Report of the Mount Sinai Hospital of the City of New York* (New York: Press of Stettiner Brothers, 1910), 24, http://archive.org/details/annualreport1907nnoun.
[113]Mount Sinai Hospital, *Fifty-eighth Annual Report of the Mount Sinai Hospital of the City of New York* (New York: Press of Stettiner Brothers, 1911), 22, https://archive.org/details/annualreport1910moun.
[114]Mount Sinai Hospital, *Fifty-ninth Annual Report of the Mount Sinai Hospital of the City of New York* (New York: Press of Stettiner Brothers, 1912), 20, https://archive.org/details/annualreport1910moun.
[115]Mount Sinai Hospital, *Sixtieth Annual Report*, 19.
[116]Mount Sinai Hospital, *Fifty-fifth Annual Report of the Mount Sinai Hospital of the City of New York* (New York: Press of Stettiner Brothers, 1908), 40, http://archive.org/details/annualreport1907nnoun.
[117] Mount Sinai Hospital, *Fifty-ninth Annual Report*, 23.

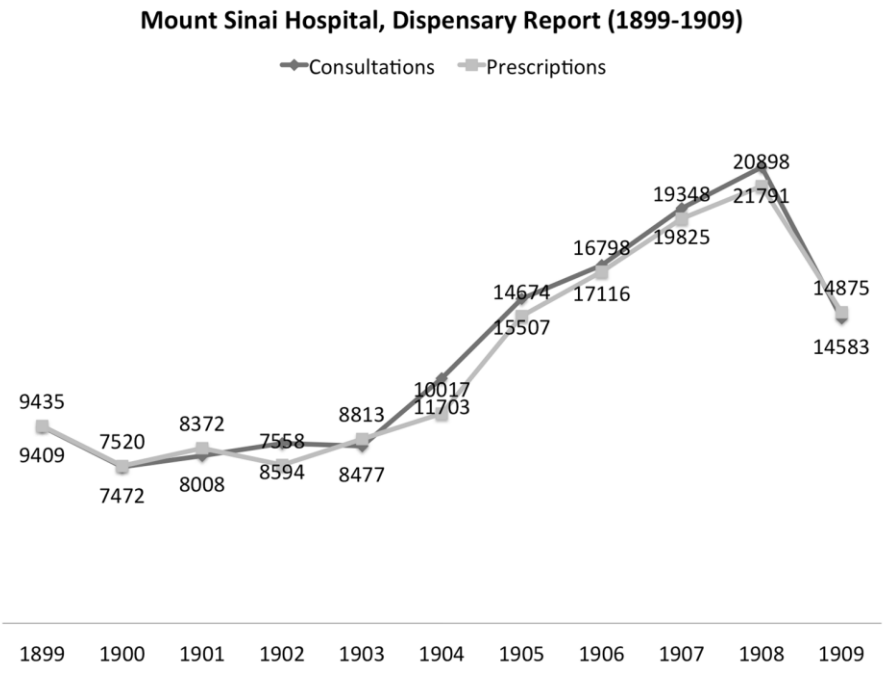


Figure 1

Source: Dispensary Reports, in *Annual Reports of Mount Sinai Hospital*, 1899-1909.

In addition to Jewish hospitals, there were nurses who helped the poor through home visitations. In 1913, the proportion of patients treated at home was very large, as the *New York Times* reported: "[i]nvestigations on the lower east side indicate that there were about 145,236 cases of sickness in this section during 1910, or 234 per 1000 population. Of these 89.4 per cent were treated in the home and 10.6 per cent in institutions."[118] As the majority of patients, including sick children, were treated outside of hospitals, nurses providing outpatient services were very important. The Mount Sinai Training School for Nurses was proud that it trained many nurses who would go on home visitations to help the ailing children and their families, as it declared that "[w]hat nobler and higher charity can there be than the services which the trained nurse renders in the restoration of the bread-winner to his family, or of children to their parents, by means of the education and training which this Institution furnishes?"[119] The Training School noted that there were indeed a lot of challenges in treating sick children, because "their minds are not enough matured to meet with patience and resignation the suffering which attends their condition."[120] At the time, with "frequent complaint by the family who are excluded from the sick-room, that the nurses do not appreciate the condition of the children," the Training School decided that it

[118] "Health Centers to Prevent Illnesses," *New York Times*, Dec 19, 1913. ProQuest Historical Newspapers.
[119] Mount Sinai Hospital, *Report of the Mount Sinai Training School for Nurses, 1896-1897* (New York: Press of Stettiner Brothers, 1898), 11, https://archive.org/details/reportofmountsin1896moun.
[120] Mount Sinai Hospital, *Report*, 24.

was worthwhile to pay the greatest attention to foster nurses who were patient and understood children.[121]

Moreover, in Jewish neighborhoods, maternity care saved children's lives because it improved mothers' health, cared for vulnerable infants, and treated sick children. In the late nineteenth century, maternity aid on the Lower East Side was provided by the New York Lying-In Hospital. It had an outdoor service department that sent physicians and medical students to homes of the patients.[122] Cooperation with Jewish charities improved the efficiency of the hospital's service: the Out-door Department cooperated with the Jewish Relief Society, and in many cases, the latter provided bed clothing, garments, food, cash, milk and grocery supplies, and daily cleaning services to poor Jewish families.[123] These vital necessities saved impoverished children from malnutrition and ensured that they had proper clothes and clean homes. As people recognized the need for specialized maternity care, supporters opened the Jewish Maternity Home in 1906, and amalgamated with Beth Israel Hospital shortly after 1909.[124] These hospitals, which took care of mothers, newborns and children, were vital for disease prevention and control among these groups.

Intergroup Relations: A Mixed Feeling

Jewish immigrants from Eastern Europe shared religious similarities traditions with assimilated Jews, yet their cultural and economic statuses were vastly different. The relations between these two groups explained how and why acculturated Jews extended a helping hand. At first, many American Jews were unwilling to accept these new immigrants. They were worried that these underprivileged immigrants "might conceivably remain permanent wards of the charitable societies, depleting resources, filling the charitable institutions, bringing a black name to the record of American Jewry, and perhaps even inciting anti-Semitism."[125] In general, American Jews believed that sick new immigrants could barely contribute to the society. As Augustus S. Levey of the New York Hebrew Aid Society told the Paris Alliance, "[w]e, as a society, and as American citizens, can not and will not be parties to the infliction upon our community of a class of emigrants, whose only destiny is the hospital, the infirmary, or perhaps the workhouse."[126]

These attitudes gradually shifted as assimilated Jews realized that only by helping these less acculturated co-religionists could Jews lessen the negative impacts of anti-Semitism. An address delivered in 1909, at the opening of the Constituent Convention of the Jewish Community of New York City, advocated that, "[i]f we organize the Jewish Community of New York City, [… i]t will wipe out invidious distinctions between East European and West European, foreigner and native, 'downtown' Jew, 'uptown' Jew, rich and poor, and it will make us realize that the Jews are one people with a common history and with common

---

[121] Mount Sinai Hospital, *Report*, 24.

[122] A.J. Rongy, "Half a Century of Jewish Medical Activities in New York City," *Medical Leaves* 1 (1937): 156.

[123] The Society of the Lying-in Hospital of the City of New York, *Annual Report, One Hundred and Eighth Year*. (New York: Order of the Board of Governors, 1906), 54-61, https://archive.org/details/annualsoci0613soci. The Society of the Lying-in Hospital of the City of New York, *Annual Report, One Hundred and Sixteenth Year* (New York: Order of the Board of Governors, 1914), 59-67, http://archive.org/details/annualsoci1421soci.

[124] Rongy, "Half a Century of Jewish Medical Activities": 161.

[125] Mandel, "Attitude of the American Jewish Community": 28.

[126] Zosa Szajkowski, "The Attitude of American Jews to East European Jewish Immigration (1881-1893)," *American Jewish Historical Quarterly* XL, no. 3 (1951): 221-233, as cited by Lindenthal, "'Abi Gezunt'": 429.

hopes."[127] In 1912, Stanley Bero, the National Organizer of the Hebrew Sheltering and Immigrant Aid Society, appealed to Americanized Jews, encouraging them to undertake "a fuller realization of their duties toward the immigrant. It is now generally believed that the immigrant must be given intelligent guidance in order that he may become a desirable factor in our body politic."[128]

From the new immigrants' point of view, receiving help from established Jews provoked mixed feelings. On the one hand, they indeed needed and cherished help as they arrived in a foreign land; on the other hand, they felt that they were looked down upon by many of those people involved in German Jews' charitable societies. According to a Yiddish article in 1884, charities operated by German Jews, "in their beautiful offices, desks, all decorated, but strict and angry faces were so arrogant that "[e]very poor man is questioned like a criminal, is looked down upon; every unfortunate suffers self-degradation and shivers like a leaf, just as if he were standing before a Russian official."[129] Although the majority of Americanized Jews tended to show a sense of superiority, and thus the "increasingly professionalized – and often patronizing – efforts of German American Jews were often resented by the East Europeans," no one could deny that their charities were helpful as "they provided important resources and created some degree of solidarity between the two groups."[130]

Conclusion

From 1880 to 1914, the ethnic, economic, and environmental factors that influenced the health of children living in urban areas varied. Indeed, socioeconomic status often had a determining effect, as wealthy people certainly had better access to decent living conditions and medical treatments. Yet it still made a difference when ethnic groups held fast to their own religious or cultural practices, which influenced the way people organized their own lives and socialized with their communities, and thus affected their health conditions. To some extent, these practices could counteract socioeconomic factors. In the case of East European Jews, their low alcoholism rate had a positive influence on their children's health. Beyond this, thanks to Jewish charities and hospitals operated by American Jews, new immigrant families received charitable assistance and were able to obtain good medical resources; therefore, the Jewish children remained relatively healthy despite their poor living conditions.

For most immigrant children in New York City, their health was exposed to risks of insufficient nutrition, inadequate sanitation, and lack of medical care. Many assimilated Jews offered generous help and one by one tackled these problems. Charitable societies gave out daily necessities to make sure that children born in poor families had enough food and clothing. The charities worked with Jewish hospitals, nursing services, and maternity homes. These medical institutions were open to the poor and the needy, and they became larger and more professional to cater to the growing number of East European immigrant children.

---

[127] Judah L. Magnes, *The Jewish community of New York City, 1877-1948* (New York, 1909), 12-13, http://hdl.handle.net/2027/loc.ark:/13960/t19k4pk80.

[128] Stanley Bero, *Chart and Explanatory Outline Regarding 81 Cities in 24 States With Particular Reference to the Jewish Communal Development* (1912), 6, http://www.archive.org/details/chartexplanatory00bero.

[129] Arthur A. Goren, "New York City," *Encyclopaedia Judaica* 12, 1971: 1092, as cited by Berger, "American Jewish Communal Service 1776-1976": 232-3.

[130] Norwood, ed., "German Jews in America," *Encyclopedia of American Jewish History*, 35.

Although the relationship between Americanized immigrants and new immigrants was not always harmonious in the beginning, these two groups, in general, actually drew closer because of their shared culture and shared challenge – anti-Semitism. The government's neglectful attitude towards challenges of hygiene on the Lower East Side made established Jews feel a responsibility to help new immigrants, and thus reinforced these intergroup relations. Although there were some tensions caused by the contemptuous manner that Americanized Jews sometimes demonstrated, there was no denying that generous help from their assimilated counterparts was crucial for new immigrants as they entered a foreign land and endeavored to make a successful transition to their new lives.

References

Adler, Cyrus ed. *The American Jewish Year Book*. Philadelphia: The Jewish Publication Society of America, 1899. http://www.ajcarchives.org/AJC_DATA/Files/1899_1900_5_LocalOrgs.pdf.

Bender, Daniel. "'A Hero... for the Weak': Work, Consumption, and the Enfeebled Jewish Worker, 1881-1924." *International Labor and Working-Class History* 56 (1999): 1-22. http://www.jstor.org/stable/27672593.

Berger, Graenum. "American Jewish Communal Service 1776-1976: From Traditional Self-Help to Increasing Dependence on Government Support." *Jewish Social Studies* 38, no.3/4 (1976): 225-246. http://www.jstor.org/stable/4466936.

Bero, Stanley. *Chart and Explanatory Outline Regarding 81 Cities in 24 States With Particular Reference to the Jewish Communal Development*. 1912. http://www.archive.org/details/chartexplanatory00bero.

Board of Health of the City of New York. *The Action of the Health Department in Relation to Pulmonary Tuberculosis and the Scope and Purpose of the Measures Recently Adopted for its Prevention*. Albany and New York: Wynkoop Hallenbeck Crawford Co., 1897. Google Book.

Carr, John Foster. *Guide to the United States for the Jewish Immigrants: A Nearly Literal Translation of the Second Yiddish Edition*. New York: The Connecticut Daughters of the American Revolution, 1912. Google Book.

The City of New York Department of Street Cleaning. *Annual Report.* 1915: 12, http://archive.org/cletails/annualreportofde00newy_5.

Cohen, Naomi W. *Encounter with Emancipation: The German Jews in the United States 1830–1914*. Philadelphia: The Jewish Publication Society of America, 1984. EBook, Varda Graphics, Inc.

Condran, Gretchen A., and Ellen A. Kramarow."Child Mortality among Jewish Immigrants to the United States." *The Journal of Interdisciplinary History* 22 No. 2 (1991): 223-254. http://www.jstor.org/stable/205867.

"The Cost of Living," *Los Angeles Times*, May 13, 1917, 12.ProQuest Historical Newspapers.

Dwork, Deborah. "Health Conditions of Immigrant Jews on the Lower East Side of New York: 1880-1914." *Medical History* 25, no.1 (1981): 1-40. doi: 10.1017/S0025727300034086.

Feld, Marjorie N. *Lillian Wald: A Biography*. Chapel Hill: The University of North Carolina Press, 2008. Kindle Edition.

Fishberg, Maurice. "Health and Sanitation of the Immigrant Jewish Population of New York." *The Menorah* 33, no. 2 (1902): 72-82. Digitized by Google.

Fox, Bertha. "The Movies Pale in Comparison." in *My Future Is in America: Autobiographies of Eastern European Jewish Immigrants*, ed. Jocelyn Cohen et al. New York and London: New York University Press. ProQuest ebrary.

"Health Centers to Prevent Illnesses," *New York Times*, Dec 19, 1913. ProQuest Historical Newspapers.

Henry, Arthur. "Among the Immigrants." *Scribner's Magazine*, January – June, 1901, no. 29, 301-311. https://babel.hathitrust.org/cgi/pt?id=mdp.39015030597127;view=1up;seq=5.

Hersch, Liebman. "International Migration of the Jews." *International Migrations*. New York, 2 (1931): 471-520. http://www.nber.org/chapters/c5117.

Kraut, A. M. and D. A. Kraut. *Covenant of Care: Newark Beth Israel and the Jewish Hospital in America.* New Brunswick: Rutgers University Press, 2006. ProQuest ebrary.

Irving A. Mandel, "Attitude of the American Jewish Community Toward East-European Immigration: As Reflected in the Anglo-Jewish Press (1880-1890)," *American Jewish Archives* (1950): 11-36. http://americanjewisharchives.org/publications/journal/PDF/1950_03_01_00_mandel.pdf.

Klein, Felix. *In the Land of the Strenuous Life*. Chicago: A. C. McClurg & Co., 1905. Google Book.

Kraut, Alan M. "Healers and Strangers: Immigrant Attitudes toward the Physician in America – A Relationship in Historical Perspective." *JAMA* 263, no. 13 (1990): 1807-1811. doi: 10.1001/jama.1990.03440130095032.

Lindenthal, Jacob J. "'Abi Gezunt': Health and the Eastern European Jewish Immigrant." *American Jewish History* 70 no. 4 (1981): 420-441. http://www.jstor.org/stable/23881910.

Magnes, Judah L. *The Jewish community of New York City, 1877-1948.* New York, 1909. http://hdl.handle.net/2027/loc.ark:/13960/t19k4pk80.

Markel, H., and A. M. Stern. "The Foreignness of Germs: The Persistent Association of Immigrants and Disease in American Society." *The Milbank Quarterly* 80, no. 4 (2002): 757-788. http://www.jstor.org/stable/3350445.

Markel, H., and A. M. Stern. "Which Face? Whose Nation? Immigration, Public Health, and the Construction of Disease at America's Ports and 1891-1928." *American Behavioral Scientist* 42 No. 9 (1999): 1313-31. doi: 10.1177/0002764299042009010.

"Mayor Acts to Stop Peril from Refuse," *New York Times*, June 29, 1907, 1. ProQuest Historical Newspapers.

Mount Sinai Hospital. *Fifty-fifth Annual Report of the Mount Sinai Hospital of the City of New York.* New York: Press of Stettiner Brothers, 1898, 1900, 1908, 1910-13. https://archive.org/details/reportofmountsin1896moun.
http://archive.org/details/annualreport1899moun.
http://archive.org/details/annualreport1907nnoun.
https://archive.org/details/annualreport1910moun.

Norwood, Stephen H., and Eunice G. Pollack, ed. *Encyclopedia of American Jewish History*. Santa Barbara: ABC-CLIO, Inc., 2008. EBook, World Wide Web.

Riis, Jacob A. *How the Other Half Lives: Studies Among the Tenements of New York.* New York: Trow's Printing and Bookbinding Company, 1890; Project Gutenberg, 2014. http://www.gutenberg.org/ebooks/45502.

Riis, Jacob A. *The Children of the Poor*. New York: Charles Scribner's Sons, 1892; Gutenberg Project, 2010. http://www.gutenberg.org/ebooks/32609.

Romanofsky, Peter. "'To Save... Their Souls': The Care of Dependent Jewish Children in New York City, 1900-1905." *Jewish Social Studies* 36, no. 3/4 (1974): 253-261. http://www.jstor.org/stable/4466835.

Rongy, A.J. "Half a Century of Jewish Medical Activities in New York City." *Medical Leaves* 1 (1937): 151-163.

The Society of the Lying-in Hospital of the City of New York, *Annual Report, One Hundred and Eighth Year.* New York: Order of the Board of Governors, 1906. https://archive.org/details/annualsoci0613soci.

The Society of the Lying-in Hospital of the City of New York, *Annual Report, One Hundred and Sixteenth Year*. New York: Order of the Board of Governors, 1914. http://archive.org/details/annualsoci1421soci.

"Some Neglected Streets: Complaint of the Condition of Thoroughfares on the Lower East Side," *New York Times*, September 1, 1901, 8. ProQuest Historical Newspapers.

University Settlement Society. *Report of the Year's Work*. New York: Concord Printing Co., 1894. http://archive.org/details/reportuniversityOOuniv.

Wenger, Beth S. "Jewish Women of the Club: The Changing Public Role of Atlanta's Jewish Women (1870-1930)." *American Jewish History* 76, no. 3 (1987): 311-333. http://www.jstor.org/stable/23883223.

Worth, Richard. *Immigration to the United States: Jewish Immigrants.* New York: Fact on Files, Inc., 2005. Ebook, World Wide Web.

Potential for Developers and Investors in Diabetes Apps to Profit by Improving the Chinese Healthcare Industry

Cheng XU

Author background: Cheng XU grew up in China and currently attends WHBC of Wuhan Foreign Languages School, located in Wuhan, China. Her Pioneer seminar topic was in the field of business and titled "How the Business Models of Emerging Apps Can Improve Healthcare."

Abstract

This paper aims to show investors and app developers how to explore the potential of the Chinese diabetes industry by selling diabetes apps. It starts by describing the extent of disease in China and the jobs-to-be-done for patients. It analyses the business model used by the Chinese healthcare system to treat the disease. It then identifies the gap between patients' needs and what the current Chinese diabetes industry provides. In the second part, the paper reviews 40 current diabetes apps. It uses the method outlined in Mark Johnson's business model framework in order to evaluate how each app supports diabetes disease management. It then concludes that though hundreds of innovative apps in China today offer to help patients manage the deadly disease, it seems that most of them demonstrate limited functionality because the business models are not sustainable. Finally, the paper shows another gap between what patients need and what current apps provide. This is exactly the place where investors and app developers in the Chinese diabetes industry could gain huge profits. It then gives investors and developers suggestions on how to improve their business models to lower costs and create higher standards of care for sufferers of diabetes in China. Overall, filling the gap in the Chinese diabetes industry is an excellent opportunity for investors and developers.

Current Chinese Healthcare Industry

The problem of diabetes has become increasingly severe in China in the last few decades. According to one IDF (International Diabetes Federation) Diabetes Atlas report, 96 million diabetes patients suffer in China, which accounts for nearly 9% of the population.[i] In other words, 1 in 10 Chinese adults suffers from diabetes. To make matters worse, this report still underestimates the data. Researchers involved in IDF's report suggest there are 493.4 million people with pre-diabetes who have abnormally high blood glucose levels[ii]. The number of potential patients is dramatically increasing. China has become the country with the largest number of people suffering from diabetes, which also makes it a country with the highest percentage of diabetes patients of the total world population.[iii] In 2015, China spent 173 billion yuan in diabetes treatment, which accounts for 13% of national healthcare expenditure.[iv] The overall diabetes problem in China is alarming.

Diabetes is a chronic disease with deferred consequences that can only be managed with extensive behavioral changes. It is caused when the pancreas fails to produce insulin due to the overly high blood sugar level. This deadly disease can make patients suffer from nerve

damage, amputation, retinal damage, cardiovascular disease and so on due to mismanagement of the disease. As a result, to effectively manage the disease on a daily basis is very important but not easy for patients. This is because patients lack a system that can manage their daily treatment as a whole. At this point, Christensen's concept of the job-to-be-done precisely describes patients' needs for reliable disease management.[v] Job-to-be-done for patients means "the basic, root problem they need to solve, or the result they want to achieve."[vi] An NIH (United States National Institute of Health) report which teaches diabetes patients how to live with diabetes, identifies and summarizes 12 crucial jobs-to-be-done for patients:[vii]

Risk calculator	For individuals who may have diabetes, "diabetes prevention is proven, possible and powerful."[viii]Those people need effective tools to calculate the possibility of getting diabetes at any time.
Basic knowledge	To take diabetes seriously and understand what happens to them, patients would like to learn about every aspect of diabetes. They want a well-organized, readable and reliable summary of all essential and basic information about diabetes.
News update	It is helpful for patients to keep informed of the newest reports about diabetes such as medication and treatment. In this way, they can have more choices for disease management.
Identify symptoms	Diabetes patients need to check their feet every day to see if there are cuts, blisters, red spots or swelling. As soon as they identify these symptoms, further treatment is immediately needed. It is helpful for patients to have a list summarizing all possible symptoms and consequences.
Blood and glucose tracking	For diabetes patients, keeping track of their blood sugar level matters. They need to check it one or more times a day and keep a record. It is helpful if they can be reminded to test the blood level and do not have to make the record themselves.
Nutrition tracking	Diabetes patients should eat low calorie food that is high in fiber. They need also to manage portion sizes of meat, vegetables, grain, etc. As a result, skills to make suitable and healthy meal plans are critical for patients.
Fitness tracking	Diabetes patients should be more active to stay at or get to a healthy weight level. Also, it is better for them to increase muscle strength through exercise. To better adhere to the exercise plan, they need reminders, instructions and an exercise record.
Social networking	Mental health is important for diabetes patients because stress can further aggravate the disease. It is helpful if they find others to listen to their concerns, allowing them tofeel better. Also, they can seek peer support in terms of tips, instructions and mental support.
Medication instructions	Diabetes patients must take medicine every day. Changes in their blood glucose level mean they need to adjust their insulin dosage according to a professional reference and also be more attuned to potential side effects. They need professional and authoritative references providing information about how to manage their insulin in case they cannot get to their doctors immediately or often enough.

Medication consumption	In addition to adjusting insulin levels, patients also pay attention to buying medication at the right price. They would like to be able to compare types and prices of medicine in different pharmacies.
Consulting with doctors	Diabetes management is a highly personalized process that patients need to do mostly by themselves. Therefore, they need quick and convenient connections with doctors in case they have any questions. Online connections which can transfer images, audio and other required information are efficient because they can directly describe and show their symptoms to doctors.
Patients' conditions summary to doctors	Diabetes patients need to record a large amount of information about their health conditions. Not only is this record essential for patients to monitor their health conditions, but it is also useful for doctors to design more effective treatment plans. So patients benefit from an easy method of recording nutrition, exercise, blood glucose and other key data and from the capability to report any changes in their health record to their doctors. It would be especially beneficial if the patients were able to provide this information to the doctor periodically rather than just at the time of an office visit.

In order to help patients effectively manage their disease, it is crucial for the current Chinese diabetes industry to meet these jobs-to-be-done for patients completely and in an organized way. However, the current Chinese healthcare system, which includes mostly public general hospitals, cannot fully meet the critical jobs-to-be-done in a cost-effective way.[ix] In other words, there is a gap between what the current Chinese healthcare system provides and what patients need urgently. Not only does this gap increase the cost of the whole Chinese healthcare industry but it also has additional negative consequences on patients' health outcomes. Table 1 (see end of the paper, above endnotes) represents the assessment of the extent to which the current Chinese healthcare industry meets patients' needs.

As Table 1 (see end of the paper, above endnotes) shows, the current Chinese diabetes industry does not offer a practical and cost-effective way to meet patients' jobs-to-be-done. Figure 1 (see end of the paper, above endnotes) summarizes the current Chinese diabetes industry's problematic business model adapted from Johnson's business model framework.[x]

The Chinese diabetes business model needs to improve on every level. Otherwise, it has three negative impacts on 96 million Chinese diabetics. First, the Chinese general hospitals are too large to focus on the specific needs of diabetes patients. As a result, patients cannot get effective treatment and the cost keeps increasing. Second, the way doctors get paid is the opposite of the patients' interests. Ideally, patients with a chronic disease such as diabetes manage their disease effectively and have fewer complications. Yet, the revenue model suggests that doctors and hospitals earn more if patients come to hospitals more frequently due to complications. Third, the frequency with which patients need to go to hospitals to care for symptoms such as nerve damage demonstrate that the healthcare resources and processes do not now provide an effective connection between doctors and patients. The majority of costs are incurred by those unnecessary processes. To conclude, the problematic business model needs improvement and innovation.

The current business model of the Chinese diabetes industry falls short. This shortcoming means too many Chinese people suffer nerve damage, amputations and even death. The business model to manage diabetes needs improvement. The business model should have a consumer value proposition focused only on chronic diseases, or more specifically on diabetes. Also, doctors should get paid according to patients' wellness, not illness. In terms of processes and resources, hospitals should explore the full potential of the internet and other technologies patients are familiar with such as applications on smart phone. In this way, they could achieve the goal of providing patients with reliable and convenient disease management services. Also, this kind of service should be integrated rather than divided into separate parts. This disruption of the current business model of the Chinese healthcare industry is the key to changing the status quo. Now apps on smartphones have the potential to serve as this disruptor, providing patients with better disease management services.

Analysis of Apps Available in Market

Many researchers have studied the general impact of diabetes apps on patients' diabetes management and the diabetes industry. Basically, those researchers have covered the trend of diabetes apps development, their functionality, and the gap between expectation and reality. Most analyses are based on apps in the English language.

Generally, the number of diabetes apps available in app stores is increasing. According to one study, in July 2009, there were 60 diabetes apps on iTunes for the iPhone.[xi] By February 2011, the number of diabetes apps available increased by 400% to 260. This increase in app usage demonstrates the increasing number and improved attitude of patients, more of whom choose to use apps to manage their diabetes symptoms. The increasing number of apps shows patients' approval, enriched functionality, and potential in this market.[xii]

Many researchers conduct systematic reviews of the functionality of apps found in the market but they cannot find a comprehensive app. One study analyzed 492 diabetes apps available in the iPhone app store.[xiii] It discovered that 33% of the total apps exist for health tracking; 22% of the total apps exist for teaching and educational purposes; 8% of the total apps exist for food reference; 5% of the total apps exist for social networking. Also 8% of the total apps are physician directed. Similarly, Demidowich analyzes 42 unique diabetes apps in the Android Market.[xiv] The result is that 36 apps, 86% of total apps, exist for self-monitoring blood glucose; 19 apps, 45% of the total apps, can help patients with diabetes medication; 11 apps, 26% of the total apps, can calculate prandial insulin dosage. Two studies show that a majority of diabetes apps only perform basic tracking functions. Also, regardless of the operating system, few of those diabetes apps tested provide comprehensive services for patients to self-manage their disease.[xv]

Existing research shows that a gap exists between what evidence-based recommendations for patients require and what functionalities the apps offer. Apart from problems associated with functions, one study demonstrates that patients have safety concerns with the use of those diabetes apps because there is not a third party review body to assess the apps for patients.[xvi] App developers themselves currently do this job. Also, privacy and other information transmitted by patients are not secured by any organizations or governments. Another study emphasizes the importance of the integration of apps with doctors and health records. These current drawbacks of diabetes apps point to the potential of the market to improve.[xvii]

This literature review provides a general overview of diabetes apps. The research covers a wide range of topics. They include the growth trend of diabetes apps, those apps' imperfect functionality and current drawbacks. At this point, this paper narrows the topic down to the situation of the Chinese diabetes market in particular. It aims to provide specific insight into the Chinese diabetes market.

In terms of data sources, I researched apps using the two most commonly used Chinese app stores. These platforms are apk.91.com and wandoujia.com. I did my research on these two websites in June 2016, using the key word "Diabetes" in Chinese. In this study, I used some inclusion and exclusion criteria to choose 40 typical apps. First, I focused on apps supported by Android to avoid repetition. The operating platforms are not a critical issue since the result would be the same if I had chosen apps supporting IOS. Second, I included apps only focusing on diabetes, excluding some apps which also target other chronic diseases at the same time such as hypertension. These apps diversify on too many areas at the same time, which makes them less than ideal for this research. Third, I included only apps designed for patients. This essay aims to show how to meet patients' jobs-to-be-done, not doctors' or hospitals' jobs-to-be-done. Fourth, I only chose Chinese apps developed by Chinese companies. In this way, I can present the overview of the Chinese diabetes industry in particular. Fifth, I only included a small number of apps which focus exclusively on references and information. Strictly informational apps have too simple functions. They are not qualified as effective self-management apps. By all the criteria above, I chose 40 apps to form a table to analyze and assess them.

After analyzing the functions of 40 apps, I used Booz balls, demonstrated in Table 2 (see end of the paper, above endnotes), to present assessment of their overall performance of meeting patients' jobs-to-be-done. The assessment is based on how many jobs-to-be-done for patients the app meets and to what extent the app meets those jobs-to-be-done.

Table 3 (see end of the paper, above endnotes) presents 40 apps chosen by the method mentioned previously with analysis of each. Apart from the name and company of each app, each column represents a job-to-be-done of diabetes patients summarized in section 1. The final column serves as an overall assessment of each app by booze ball. Each row shows one of the 40 apps.

This essay closely looks at 4 typical apps ranged from "Completely" if they cover all jobs-to-be-done, to "Incompletely" if they are missing some components to clarify the way apps work and the standard of evaluation. They are "微糖", "血糖记录", "甜蜜糖尿病" and "糖尿病护士". These apps have different sets of functions and can represent the general way apps work.

"微糖" is a free diabetes app which supports 11 out of 14 jobs-to-be-done for diabetes patients. The relative comprehensive functions help patients self-manage the disease conveniently with only one app. This app provides the required information and references for diabetes patients. In terms of self-management, it includes all three necessary daily records and provides a social platform for patients to find similar people to get motivation. Its function of connection with doctors can increase treatment efficiency and save cost. Something that makes this app stand out and get a high score is its partnership with general hospitals. This partnership gives the app certification of quality on the psychological level to

gain patients' trust, which means it can bring potential profit to the app developers. One drawback of this app is that it does not have a link with the government. The lack of government support means less potential to be used widely because governments can distribute the app to public general hospitals. Overall, this app locates at the top of the pyramid of current diabetes apps.

"血糖记录" is a free diabetes app which provides almost none of the jobs-to-be-done for diabetes patients, scoring only 1 out of 14. This app has a limited function, which is only blood glucose tracking. It is usually not practical for patients to download several apps together to manage the disease. Patients would not choose this app, especially if they can have other apps with more functions. As a result, this app can have a very limited positive impact on patients' self-management of the disease.

"甜蜜糖尿病" is a free diabetes app which meets 3 out of 14 jobs-to-be-done for diabetes patients. Its main functions are social networking and blood glucose tracking. Patients can share experiences and also purchase medication and devices directly from other people using this app. However, patients cannot share useful daily records and ask related questions through isolated social networking without other functions. As a result, patients cannot get motivation to adhere to treatment from this social networking. After all, social networking is auxiliary to self-management functions. Overall, this app which focuses on one main function, cannot effectively meet patients' needs.

"糖尿病护士" is a free app which supports 7 out of 14 jobs-to-be-done for diabetes patients. Patients can obtain basic information and new updates about diabetes using this app. This app also provides medication instructions. The most effective part of this app is the connection with doctors and hospitals. Patients can consult with doctors on a one-to-one basis, so doctors can have a more complete overview of patients' condition. Because it is partnering with hospitals to get contact with those doctors, this certification of quality can convince patients to use this app. In terms of daily recording functions, it only has a blood tracking function. Doctors can obtain the data conveniently. However, patients cannot efficiently self-manage their disease if fitness tracking and nutrition tracking are missing. Overall, the connection with doctors and hospitals makes this app effective, but it can be further improved by adding more daily recording functions.

There are some general patterns about Chinese diabetes apps emerging from table 3. Looking at the table, this essay broadly summarizes eight findings:

- None of the 40 apps provide all the functionality requirements. In other words, none of these 40 apps include all the functions required to fulfill jobs-to-be-done for diabetes patients.
- All 40 apps are free. There are no sustainable revenue models like subscription or data selling models, for example.
- 28 of the 40 apps have at least one basic daily recording function. Those functions include blood glucose tracking, fitness tracking and nutrition tracking.
- 3 of the 40 apps have links with hospitals, government or both.
- 19 of the 40 apps have at least one function in terms of connection with doctors.

These functions include patients' health conditions summary sent to doctors and consulting with doctors.

- 4 of the 40 apps have a risk calculator function. In other words, only a few apps can provide potential patients with diabetes prevention.
- 12 of the 40 apps have a social networking function. However, social networking is crucial for patients' motivation to adhere to treatment.
- 7 of the 40 apps tested have a medication consumption function. None of them have links with pharmaceutical companies. As a result, the effectiveness of these apps to meet patients' medication consumption needs is uncertain.

To conclude, the Chinese diabetes apps market is immature and needs improvement. First, none of the apps tested can perform a comprehensive set of functions to meet the complete patients' jobs-to-be-done. In other words, none of the current diabetes apps can help patients conveniently self-manage their disease. It is not practical for patients to use several apps together. Second, none of the apps tested have a sustainable revenue model which is crucial for these apps' future development. Third, only a few of these apps have links with hospitals and government. This lack of connection can result in patients' distrust, limited usage rate and no integration with patients' health records. Reasons include lack of quality certification, promotion support and uniform system with hospitals. From an analysis of table 3 (see end of the paper, above endnotes), diabetes apps can meet jobs-to-be-done for diabetes patients. However, a gap exists between what patients require to effectively manage their disease and what current apps actually provide.

Expectations of Promising Future Chinese Diabetes App Industry

The current business model of the Chinese healthcare industry falls short. This business model of general hospitals is not suitable for the treatment of chronic diseases like diabetes. The inefficiency allows 96 million diabetes patients to suffer from complications and increasing costs. The gap between what Chinese general hospitals provide and what patients require to manage their disease can be largely filled by diabetes apps.

However, a gap also exists between what patients can expect from ideal apps and what current Chinese diabetes apps can provide. The problems this essay identifies in section 2 are all drawbacks of current Chinese diabetes apps. Yet, they also point to areas where current Chinese diabetes apps can improve. Clearly, the potential of a Chinese diabetes app industry exists. The opportunity for investors and apps developers to gain huge profits from a Chinese diabetes app industry comes from the gap between ideal apps and current apps.

In order to seize the opportunity, investors and app developers could aim at designing effective diabetes apps for Chinese diabetes patients. In other words, their goal could be to make the app meet all of patients' jobs-to-be-done to the largest extent and have stable connections with the current healthcare system. At this point, the app companies will not be selling just apps. Instead, they can launch a set of diabetes management services with apps as the method by which patients access a set of relevant services. With this service, patients with diabetes can effectively self-manage every aspect of their disease and have immediate contact with professionals. In order to meet this expectation, app developers could build a brand new business model focusing on diabetes services, not merely the apps themselves.

For every new app, it is useful to adapt MVP (minimum variable product) strategy, which means to start with basic functions and then to get feedback from early consumers.[xviii] In other words, in the first step of launching a successful app, app developers could provide some necessary and basic functions in the app and wait for the feedback of early patients. This strategy is practical because those feedbacks can provide app developers with information on what consumers need and therefore lead the app design to the right track in terms of further upgraded version of the app. Moreover, this strategy is cost-effective because it can prevent the app company from wasting resources on functions which turn out to be not needed by the patients. MVP strategy can render app companies thrive in a rather competitive market.

The most important and essential step is that app developers could redesign current diabetes apps in terms of their functions. An ideal diabetes app should have all the functions mentioned in Table 3 (see end of the paper, above endnotes). In other words, it would be effective and beneficial for diabetes patients to have as few apps as possible to manage their disease. Having several apps together to manage the disease can be impractical and confusing to diabetes patients, especially to older patients.

Fifth, app companies could have sustainable revenue models to ensure future development. Since diabetes management service is the ideal product app companies should launch, the revenue model can be a monthly subscription fee. That might be 50 to 100 yuan per month. It is estimated that Chinese diabetes patients have to spend 18 thousand yuan per year in terms of treatment. This healthcare expenditure is 5.1 times the amount spent by a healthy person.[xix] Chinese diabetes patients would be able and willing to afford the small cost monthly. It is reasonable and attractive for them to effectively self-manage their disease with less cost. With roughly 100 million diabetes sufferers in China, that would be a total addressable market of 120 billion yuan per year. This is the huge profit app developers can gain from the immature Chinese diabetes industry. This subscription revenue model should be shared with partners to ensure stable and sustainable connections. For example, doctors from general hospitals in second-level support help desks should get paid for answering patients' questions. Otherwise, doctors will not be motivated to do the job, and the partnership will not be secured. Also, it is possible that app developers can sell patients' data to hospitals and pharmaceutical companies for research use. All in all, it is necessary for app companies to have sustainable revenue models to seize the full potential of the Chinese diabetes app industry.

Moreover, the app company should have a sustainable revenue model to support business activities. Generally, there are two models suitable for diabetes apps selling which are freemium and in-app AD monetization.

Freemium means app company can make the core of its disease management free and charge less price-sensitive patients on upgraded and more comprehensive version. For instance, the company can make first-level support free and charge patients with second-level support mentioned previously. This revenue model is beneficial especially for diabetes app companies for three reasons. First, while paid apps block most patients, freemium can attract a large consumer base. Second, freemium itself can be a promotion. Patients can gradually know the functions and benefits of this apps through their own experience. The tendency to pay to get the full version is also ensured by their sense of value-adding. Third,

freemium can not only attract but also keep patients to form loyal consumer base. After patients are used to this app, they are not willing to change as all the key health data is in this app. However, app companies should carefully consider which part of the service should be free to guarantee the greatest value added after patients pay for the upgraded version. Moreover, this revenue model should be shared with partners to ensure stable and sustainable connections. For example, doctors from general hospitals in second-level support help desks should get paid for answering patients' questions. Otherwise, doctors will not be motivated to do the job, and the partnership will not be secured.

The potential of this revenue is huge. If patients are willing to pay for upgraded version, the fee might be 50 to 100 yuan per month. It is estimated that Chinese diabetes patients have to spend 18 thousand yuan per year in terms of treatment. This healthcare expenditure is 5.1 times the amount spent by a healthy person.[xx] Chinese diabetes patients would be able and willing to afford the small cost monthly. It is reasonable and attractive for them to effectively self-manage their disease with less cost. With roughly 100 million diabetes sufferers in China, that would be a total addressable market of 120 billion yuan per year. This is the huge profit app developers can gain from the immature Chinese diabetes industry.

Another model is allowing in-app ads, which mean app companies are paid to allow other companies' advertisements to appear in the app. This method is easy and cheap, but needs a large number of page views to attract advertisers. This essay suggests app companies use a freemium model in the beginning stage and add in-app ads when the consumer base is stable and of suitable scale. However, choosing the most appropriate revenue model largely depends on companies' individual situations. This essay also suggests companies should do market research and then decide their revenue model plans.
After app companies achieve the two steps above, they could consider the following more advanced functions according to patients' feedback. This innovative business model for diabetes app could provide patients with more effective and convenient disease management and therefore provide app companies with opportunities to gain profit.

First, app companies could create an innovative help desk service for patients because connection to professionals plays a critical role in terms of self-management. In other words, app companies could create or partner with a service that diabetes patients could call, text or email to ask questions. App companies can partner with hospitals and other service companies to deliver this service, like IT companies do with their call centers.

In terms of first-level support, app companies could create or partner with a team of trained people specializing in basic but comprehensive diabetes treatment. Nowadays, lots of treatment and symptom identification in the diabetes industry are in the area of precision medicine. Precision medicine means the provision of treatment for diseases that can be precisely diagnosed and treated by rule-based therapies, which have predictably effective consequences.[xxi] At this point, questions about the scope of precision medicine can be answered without physicians and specialists. As a result, members in the first-level support team can be trained: registered nurses or other trained people specializing in diabetes including symptom treatment and nutrition and fitness management. This first-level help desk is useful and practical for patients when their symptoms can be dealt with at home or require immediate treatment at emergency departments. It is also helpful when they have basic questions about the way to deal with their current health conditions. Dialing a nurse service provided by EvergreenHealth in the U.S. would be a suitable example in this context

as it specializes in diabetes.[xxii] EvergreenHealth is an integrated two-hospital system which provides a range of services.[xxiii] One of them is a 24-hour consulting nurse service. Patients can dial the number provided on the website to get help from trained and experienced nurses at any time if their symptoms can be treated at home or if they need immediate contact at an emergency room. This kind of service has not been available in China yet. App companies can introduce this service in their apps by creating or partnering with foreign companies.

In terms of second-level support, app companies could partner with Chinese general hospitals to back up the first-level support people. Those hospitals can create a team of doctors to answer questions in the area of intuitive medicine while making doctors' schedules available. Intuitive medicine means care for diseases which cannot be diagnosed precisely and can only be treated with therapies which have uncertain efficacy.[xxiv] When patients ask questions which people in first-level support cannot answer in the communication system mentioned above, these non-urgent questions in the area of intuitive medicine will be sent to doctors from general hospitals.

Second, app companies could also partner with general hospitals on the management level. Apart from answering questions related to diabetes management, help desks should also help patients make appointments with physicians and specialists. At the same time, hospitals can officially recommend the app directly to patients. In this way, apps can be widely used by diabetes patients through the promotion of general hospitals. Moreover, hospitals may gain more patients because patients tend to choose hospitals where they are familiar with doctors in second-level help desk support. Partnership between app companies and hospitals can be a win-win situation.

Third, app companies could partner with weight loss and fitness programs to fulfill patients' fitness needs to the largest extent. Apart from nutrition tracking, patients may be more satisfied when the app can also provide fitness instructions, personalized fitness plans and social networking. Fitness programs like Weightwatchers in the U.S. are a good example.[xxv] Weightwatchers provides over 4,000 delicious and healthy recipes. People attending this program can eat delicious and family-friendly food and still lose weight. This program also provides patients with fun and easy ideas to learn new sports. People can strengthen their health condition and lose weight without suffering. Moreover, this program will give participants personal assessments to help them gain insight into their lifestyle, goals and challenges. Therefore, participants can have focused plans to meet their specific goals. In this program, diabetes patients can have all their fitness targets met very effectively. However, the Weight Watchers' recipes and approach are designed for Americans; this essay only presents it as an example for app designers to think about. To conclude, app companies could partner with other companies providing this service to improve the range and quality of diabetes management services.

Fourth, app companies could partner with glucose monitor manufacturers. The app could be connected with patients' daily used medical device. Once patients test their glucose level, data will be automatically transferred to the app. Then the app can summarize the data and transfer to doctors periodically. In this way, patients' records will be more accurate and complete because patients do not need to record the data manually every time. Also, doctors can better assess patients' health conditions thanks to regular and accurate feedback. In this way, the app and medical device can actually promote each other like the app and hospital do. Due to the development of technology nowadays, the existence of a tiny device

implemented under patients' skin to test glucose levels every five minutes is possible. However, this vision is too bold to survive in the current Chinese diabetes industry. Chinese patients are unlikely to fully accept this advanced technology due to its potentially high price and low acceptability to unknown products especially related to healthcare.

Having the diabetes management service described above, diabetes patients can expect their diabetes management to be convenient and efficient. The benefits of this service can be fully demonstrated by the scenario that followed. Li, a middle-aged Chinese diabetes patient, uses an app that provides the diabetes management service. His typical day is like this with the help of the app: Li wakes at 7 A.M. to the ring of his iPhone. Then he hears the reminder from the app for using the medical device to check his blood glucose level. This reminder will appear throughout the day according to the changes in Li's health condition. Li uses the medical device to test his blood glucose level and the data automatically transfers to the app to calculate a summary. Now it is breakfast time. Li chooses a delicious and family-friendly breakfast recipe from thousands of those available in the app. After eating breakfast, Li starts doing morning exercises according to the personalized fitness plan he makes with the app. With a healthy body, Li can get to work with high efficiency. In the afternoon, Li also enjoys a healthy meal with the help of the app. After that, the exercise reminder signals him again. In his spare time, Li reads the latest news updates about diabetes on the app. In the afternoon, Li receives the monthly summary of his glucose level with doctors' suggestions. After reading the summary, Li has some questions and he dials the number of the help desk. Some of his difficult and quite professional questions are sent to doctors by nurses. Li has all his questions answered just by phone. At the end of the day, Li receives two fitness assessments because fitness assessments are made both at the end of each day and at the end of each month. Since Li feels that the current plan has become too easy, he reports his feeling to the app and adjusts the goal. Then Li immediately gets his new plan from the app. Li can now start a new day with his disease better managed.

To conclude, the current Chinese healthcare system for diabetes treatment is ineffective. It cannot fulfill Chinese diabetes patients' jobs-to-be-done completely and in a well-organized way. Diabetes apps, modern technology in people's daily lives, are an effective disruption for the current Chinese healthcare industry for diabetes. The immature Chinese diabetes app market represents an opportunity for investors and app developers to gain substantial profits. Also, patients suffering from poor disease management from general hospitals can benefit in terms of their health and expenditures. In order to seize this opportunity, app developers and investors should consider the vision of this essay. App companies should create a whole set of diabetes management services with apps providing access for patients to get services. This set of services should meet all patients' needs to the largest extent and have sustainable revenue models which have links with the current healthcare system. With this vision, app developers and investors can create a new 120-billion-yuan industry, save 96 million lives, and keep 493.4 million people with pre-diabetes from getting the full disease. Overall, this essay shows a clear direction for investors and app developers to explore the Chinese diabetes app industry.

Figure 1: Current Chinese Diabetes Business Model

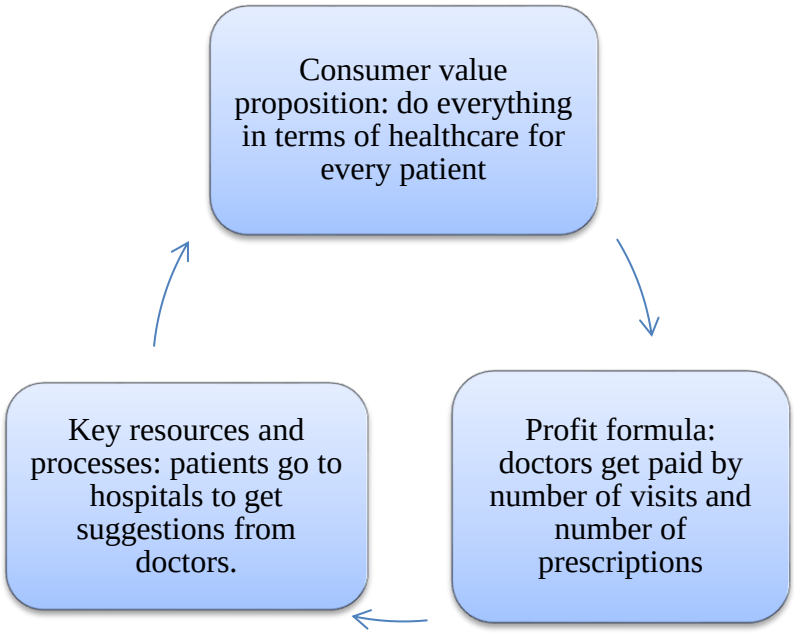


Table 1 – Assessment of jobs-to-be-done by current Chinese healthcare industry

Jobs-to-be-done for patients	**The way current healthcare industry arranges to have jobs done**	**Extent jobs are completed**	**Consequence**
Risk calculator	Government posts authoritative guidelines on line	◒	Number of potential patients and undiagnosed cases increases.
Basic knowledge	Authoritative guidelines provided by government on its official website	◓	Lacking a collection of all, patients cannot effectively identify and remember useful information.
Identify symptoms	Doctors tell patients possible symptoms during regular visits	◑	Patients may not get and remember details, creating the possibility of complications.
News update	Patients mostly rely on themselves to look up resources.	◑	Patients cannot easily get comprehensive new information.
Blood and glucose tracking	Patients track at home and report to doctors only when they go to the hospital.	◒	There is a huge gap between doctors and patients in terms of patients' health conditions. Doctors cannot effectively design

Jobs-to-be-done for patients	The way current healthcare industry arranges to have jobs done	Extent jobs are completed	Consequence
			treatment.
Fitness tracking	Patients may or may not record themselves and report only when they go to hospital.	◒	Patients are less motivated to adhere to exercise plan.
Nutrition tracking	Patients get only general guidelines from doctors.	◒	Patients cannot design healthy meal plans themselves and therefore are more likely to have complications.
Social networking	No efforts	◒	Patients mostly rely on themselves to figure out all specific processes and may have mental stress.
Medication instructions	Patients get from doctors only when they go to the hospital.	◑	Patients go to hospitals on unnecessary circumstances and may misuse medication.
Medication consumption	Patients buy it from hospitals or find a pharmacy themselves.	◓	It is difficult for patients to compare prices and find a rare type of medication.
Consulting with doctors	Patients travel far to the hospital to meet doctors.	◑	Going to hospitals unnecessarily many times increases cost and defers treatment.
Patients' conditions summary to doctors	The only record shared by patients and doctors is medical history at the hospital.	◒	Doctors cannot give effective suggestions in time to help patients.

Key: ● completely ◓ acceptably ◑ little ◒ basically not

Table 2- Booz ball evaluation standard

To what extent does this app support the management of diabetes?

- ● Completely
- ◓ Mostly
- ◑ Not much
- ◒ Not at all

Table 3- 50 Chinese diabetes apps survey

Name	Com-pany	Price	Calculator	Knowledge	Symptoms	News	Blood Glucose	Nutrition	Fitness	Networking	Medication Instructions	Medication Consumption	Patient Summary	Consulting	Hospitals	Government	Evaluation
糖尿病心天地	礼来中国	Free		✓			✓	✓	✓	✓	✓						◑
悦糖	北京悦优博迈科技有限公司	Free	✓				✓	✓	✓	✓							◑
微糖	上海格平信息科技有限公司	Free		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		◑
血糖记录	加州厦门健康管理有限公司	Free					✓										◐
血糖高管-糖尿病助手	柏云健康科技（北京）有限公司	Free	✓				✓	✓	✓			✓					◑
糖尿病护士	启名方科技（北京）有限公司	Free		✓		✓	✓				✓		✓	✓	✓		◑
掌控糖尿病	福州康为网络技术有限公司	Free		✓			✓	✓	✓	✓	✓			✓			◑
糖尿病管理	GExperts Inc	Free					✓	✓	✓				✓				◑
甜蜜糖尿病	上海欢智网络科技有限公司	Free					✓			✓		✓					◑
糖尿病健康助手	四川掌云科技有限公司	Free					✓	✓	✓	✓	✓						◑

Name	Com-pany	Price	Calculator	Knowledge	Symptoms	News	Blood Glucose	Nutrition	Fitness	Networking	Medication Instructions	Medication Consumption	Patient Summary	Consulting	Hospitals	Government	Evaluation
医随访	健安华夏（北京）科技有限公司	Free					✓	✓					✓	✓			◑
糖尿病新世界	武汉鼎森世纪科技有限公司	Free		✓		✓											◒
慢友帮	青岛联创优内信息技术	Free								✓							◑
糖尿病医生	互动在线（北京）科技有限公司	Free		✓	✓	✓											◒
糖大大	成都唐大大科技有限公司	Free				✓				✓				✓			◑
糖尿病计算器	珠海市奥美软件科技有限公司	Free	✓														◒
糖尿病防治	北京一点网聚信息技术有限公司	Free		✓	✓	✓											◑
糖尿病漫画集	珠海市奥美软件科技有限公司	Free		✓													◒
糖尿病参考值	珠海市奥美软件科技有限公司	Free		✓													◒

Name	Company	Price	Calculator	Knowledge	Symptoms	News	Blood Glucose	Nutrition	Fitness	Networking	Medication Instructions	Medication Consumption	Patient Summary	Consulting	Hospitals	Government	Evaluation
糖尿病药物速查	珠海市奥美软件科技有限公司	Free									✓						◑
糖护士-糖尿病	北京糖护科技有限公司	Free					✓	✓	✓			✓	✓				◑
糖尿病专科网	杭州物趣科技有限公司	Free		✓		✓											◒
医到病除	北京德信融通科技有限公司	Free		✓		✓	✓	✓	✓	✓			✓	✓			◑
大糖医	汇马医疗科技	Free		✓		✓	✓						✓	✓			◑
好豆菜谱	湖南红图网络科技有限公司	Free								✓							◒
糖尿病关怀	诺和关怀俱乐部	Free		✓			✓						✓	✓			◒
糖尿病百科	北京凯晟达科技有限公司	Free		✓	✓									✓			◒
安测健康	深圳市前海安测信息技术有限公司	Free		✓			✓		✓				✓	✓			◑
随糖	科学技术文献出版社	Free		✓			✓	✓	✓				✓	✓			◑
掌上糖医	杭州康晟健康管理咨询有限公司	Free		✓			✓	✓	✓	✓		✓					◑

Name	Com-pany	Price	Calculator	Knowledge	Symptoms	News	Blood Glucose	Nutrition	Fitness	Networking	Medication Instructions	Medication Consumption	Patient Summary	Consulting	Hospitals	Government	Evaluation
深敏血糖	上海深敏医疗科技有限公司	Free		✓			✓	✓			✓						◑
易糖	易联众信息技术股份有限公司	Free					✓	✓	✓					✓			◑
知糖	北京重生科技有限公司	Free					✓	✓			✓						◒
控糖卫士	北京新净科技有限公司	Free	✓	✓			✓	✓	✓			✓		✓			◑
春雨糖管家	北京春雨天下软件有限公司	Free					✓	✓	✓		✓		✓	✓		✓	◑
依聊	北京康语科技有限公司	Free					✓			✓							◒
慢病健康到家	天津天士力电子商务有限公司	Free		✓			✓					✓					◑
糖伴	上海博业信息科技有限公司	Free		✓			✓	✓	✓				✓	✓			◑
金典糖医	北京金典伟业科技有限公司	Free		✓			✓	✓	✓		✓	✓		✓			◑
糖无忧	招商信诺人寿保险有限公司	Free		✓			✓							✓			◒

References

[i]Yoo, Eva. "An Influx Of Apps Take Aim At Asia's Growing Diabetes Problem." *Technode*. TechNode Inc., 02 June 2015. Web. 24 June 2016. <http://technode.com/2015/06/02/diabetes-management-platform-gather-health-raises-us2m/>.

[ii]"Bring Research in Diabetes to Global Environments and Systems (China)." *International Diabetes Federation*. International Diabetes Federation, 2015. Web. 24 June 2016<https://www.idf.org/BRIDGES/map/china>.

[iii]"Diabetes in China." *Diabetes.co.uk the Global Diabetes Community*. Diabetes.co.uk, n.d. Web. 24 June 2016. <http://www.diabetes.co.uk/global-diabetes/diabetes-in-china.html>.

[iv]"China Healthcare Expenditure reached 173 billion in 2015." *Askci.com*. askci Corporation, 11 June 2016. Web. 24 June 2016. <http://www.askci.com/news/dxf/20160611/17282027494.shtml>.

[v] Christensen, Clayton. *The Innovator's Prescription: A Disruptive Solution for Health Care.* The United States of America: McGraw-Hill, 2009. Print.

[vi]Christensen, Clayton. *The Innovator's Prescription: A Disruptive Solution for Health Care.* The United States of America: McGraw-Hill, 2009. Print.

[vii]"4 Steps to Manage Your Diabetes for Life." *National Institution of Diabetes and Digestive and Kidney Diseases*. The National Institute of Diabetes and Digestive and Kidney Diseases., April 2014. Web. 24 June 2016. <http://www.niddk.nih.gov/health-information/health-topics/Diabetes/4-steps-manage-diabetes/Pages/publicationdetail.aspx>.

[viii]"4 Steps to Manage Your Diabetes for Life." *National Institution of Diabetes and Digestive and Kidney Diseases*. The National Institute of Diabetes and Digestive and Kidney Diseases., April 2014. Web. 24 June 2016. <http://www.niddk.nih.gov/health-information/health-topics/Diabetes/4-steps-manage-diabetes/Pages/publicationdetail.aspx>.

[ix]Hou, Bin and Jin Cui. "the status quo of current Chinese healthcare industry.Tournal of Yunan College of Traditional Chinese Medicine 8 (2006): 19-21. Cqvip.com. Web. 30 Jun. 2016. <http://www.cqvip.com/read/read.aspx?id=1000202944#>.

[x] Johnson, Mark. Seizing The White Space: Business Model Innovation for Growth and Renewal. The United States of America: Harvard Business School Publishing, 2010. Print.

[xi]Chomutare, Taridzo, et al. "Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines." *Journal of medical Internet research* 13.3 (2011): e65.

[xii]El-Gayar, Omar, et al. "Mobile applications for diabetes self-management: status and potential." *Journal of diabetes science and technology* 7.1 (2013): 247-262.

[xiii]Eng, Donna S., and Joyce M. Lee. "The promise and peril of mobile health applications for diabetes and endocrinology." *Pediatric diabetes* 14.4 (2013): 231-238.

[xiv]Demidowich, Andrew P., et al. "An evaluation of diabetes self-management applications for Android smartphones." *Journal of telemedicine and telecare* 18.4 (2012): 235-238.

[xv]Demidowich, Andrew P., et al. "An evaluation of diabetes self-management applications for Android smartphones." *Journal of telemedicine and telecare* 18.4 (2012): 235-238.

[xvi]Eng, Donna S., and Joyce M. Lee. "The promise and peril of mobile health applications for diabetes and endocrinology." *Pediatric diabetes* 14.4 (2013): 231-238.

[xvii]El-Gayar, Omar, et al. "Mobile applications for diabetes self-management: status and potential." *Journal of diabetes science and technology* 7.1 (2013): 247-262.

[xviii]NIVI. *What is the minimum variable product?* 2009. Web. 31 December. 2016. <http://venturehacks.com/articles/minimum-viable-product >.

[xix] "Chinese diabetes patients' healthcare expenditure has reached 18 thousand." *Chinanews.com.* Chinanews.com. November 2010. Web. 26 June 2016. http://www.chinanews.com/jk/2010/11-22/2673228.shtml

[xx] "Chinese diabetes patients' healthcare expenditure has reached 18 thousand." *Chinanews.com.* Chinanews.com. November 2010. Web. 26 June 2016. http://www.chinanews.com/jk/2010/11-22/2673228.shtml

[xxi]Christensen, Clayton. *The Innovator's Prescription: A Disruptive Solution for Health Care.* The United States of America: McGraw-Hill, 2009. Print.

[xxii]"24-Hour Consulting Nurse." *EvergreeHealth.* EvergreenHealth, 2016. Web. 1 July. 2016. <https://www.evergreenhealth.com/24nurse>.

[xxiii]24-Hour Consulting Nurse." *EvergreeHealth.* EvergreenHealth, 2016. Web. 1 July. 2016. <https://www.evergreenhealth.com/24nurse>.

[xxiv] Christensen, Clayton. *The Innovator's Prescription: A Disruptive Solution for Health Care.* The United States of America: McGraw-Hill, 2009. Print.

[xxv]"Weightwatchers." *Weightwatchers.*Weight Watchers International, Inc. 2016. Web. 1 July. 2016. <https://www.weightwatchers.com/us/>.