Bryant University

HONORS THESIS

Revisiting the Mind- Body Paradox: Can Brain Functioning Explain Moral Reasoning?

BY: Brianna Mahan

ADVISORS • Brian Blais

Submitted in partial fulfillment of the requirements for graduation
with honors in the Bryant University Honors Program

APRIL 2009

Table of Contents

ABSTRACT

With this paper, I attempt to explore possible neural correlates of morality. We define morality as the one part, structure, or process of the brain that could be linked to an innate ability to understand and determine right versus wrong. An understanding of right of right and wrong can provide us with a sense of guilt and empathy for an action or another person. Right and wrong will be defined through a primarily Judeo-Christian perspective, as it was the principle respondent among our questionnaire. There is a possibility for differences among other religions. For that reason, we expect the neural correlate to be flexible enough to lead to variations. Mirror neurons, or neurons with the ability to excite while watching and executing an action, will be the neural correlate I will explore. Using a combination of Jaynes theory of consciousness, Hawkins hierarchical temporal memory, and a pattern recognition associative network, I will recreate a mirror neuron network, which could represent a learning pattern which develops to classify actions as "right" or "wrong" (Jaynes, 1990, Hawkins 2005).

INTRODUCTION

Many would question why I would research two fields that tend to never get along: religion and science. Both fields overwhelmingly argue they cannot, and do not, overlap. However, I find that science and religion offer a hand to one another in many explanations of natural functions (Miller 1999). For example, science can explain why a star will collapse into black hole, but religion can offer us a scope of the beauty, enormity, and implications of this event.

With this paper, I attempt to explore possible neural correlates of morality. We define morality as the one part, or structure, or process of the brain that could be linked to an innate ability to understand right versus wrong. An understanding of right and wrong can provide us with a sense of guilt and empathy for an action or another person. The implications for this finding include possible explanations for why some are prone to killing others, why women are more emotional or guilt ridden than their male counterparts, if animals have "souls", or if morality is a luxury reserved for humans. We could work to rehabilitate people who have no sense of right or wrong, and examine the usefulness and evolutionary implications of having a set of morals, and why they exist in the first place. We could examine the role of culture in the shaping of moral actions, and how those actions become the definition of morality. I propose that mirror neurons, located in various areas of the brain, are responsible for empathetic response and understanding the value of human life, belongings, and loved ones. These neurons result in a set of values and standards, which creates a common belief system of moral actions, and in turn become reinforced though societal pressures.

A clear distinction between faith and religion needs to be created, and the determination of right versus wrong needs to be defined. Faith, for the purpose of this paper, is a single person's belief in a higher power and a moral set of standards they uphold. Religion, however, is an institution; a set of distinct, written beliefs in which most people practice their faith and follow. I will use religion to provide distinctions of morality and right versus wrong. I will not be using faith-based answers because opinions can vary greatly from person to person.

Right and wrong will be defined through a primarily Judeo-Christian perspective, as it was the principle respondent among our questionnaire. There is a possibility for differences among other religions, for that reason, we expect the neural correlate to be flexible enough to lead to variations. This measure will then be used to classify right versus wrong in experimentation and explanation of theories and facts. Using a combination of Jaynes theory of consciousness, and Hawkins hierarchical temporal memory and a pattern recognition associative network, I will recreate a mirror neuron network, which could be a starting point to learning to classify actions as "right" or "wrong".

There is a large portion of this paper about consciousness. This is for two reasons. First, in order to have a moral system, we need consciousness. We must be aware of ourselves and our actions, and how they affect others. Different models of consciousness limit consciousness to certain brain structures or processes, which in turn limit the scope of consciousness. Some models, like Jaynes, limit consciousness to humans, while others, like Hawkins, limit consciousness to animals possessing a cortex (i.e. mammals) with varying degrees of conscious thought. Secondly, in exploring properties of consciousness, we are lead to certain concepts that help us in understanding properties of morality. For example, Jaynes uses the development of language and metaphor to explain consciousness, which in turn affects how we look at the function of mirror neurons (Jaynes 1990). Hawkins, on the other hand, believes we achieve consciousness through the linking of patterns and feedback, which in turn views mirror neurons as an association of sequences (Hawkins, 2005).

LITERATURE REVIEW

Jaynes' Theory of Consciousness and Metaphors

Jaynes starts explaining consciousness by looking at what it is not. Jaynes believes there is a difference between consciousness and reactivity. We are able to react to things without being conscious of it. He also believes we cannot determine the continuity of consciousness, as we are unable to be aware of unconscious thought, as we are unconscious of it. For example, when we drive, we react to numerous variables on the road, such as other cars, wind, potholes, changes in speed, and are nearly never conscious of it. Consciousness of actions can also often get in the way of what we are doing. If you stopped to think about the specific motions your figures take the write a sentence, your handwriting will most likely change.

Jaynes outlines five distinct items that are not specific to consciousness. First, Jaynes explains that consciousness is not a copy of experience. This means that your consciousness is not a mental representation of your everyday life. There is a difference between the recall of experiences and the recognition of experiences. We are able to recognize experiences very easily, however, when asked to recall experiences, we have a hard time, and they are often slightly skewed from reality. Secondly, consciousness is not necessary for concepts. A concept is something like a "tree" or an "apple". You have never seen a tree or an apple, because there are all different types of trees, and apples, but there is no one apple. For example, there are Macintosh, Green, Yellow, Granny Smith, and many more, but no apple. However, unconsciously, we all have an understanding of what an apple is. This is the concept of an apple, and consciousness is not necessary to develop these concepts. Third, consciousness is not necessary for learning. We learn things from visual and verbal cues on a daily basis, and don't even realize it. For example, Jaynes explains an experiment in which students were told to compliment every girl who wears red. Within the week, almost every girl wore red to school, and they didn't even realize the trend. Also, when having a conversation, if you reinforce certain words or subjects, the other person will start to only talk about those subjects, or use those words repeatedly, without consciousness intending to do so. This is a form of operant conditioning. Fourth, consciousness is not necessary for thinking. Jaynes points out that our judgments often are made unconsciously. For example, when judging weights, we know which one is heavier, yet we cannot explain the process we took to reach that conclusion. Our judgments, Jaynes argues, are automatic responses depending upon the materials offered. In other words, our judgments are a result of our instruction and context. Fifth, consciousness is not necessary for reasoning. Reasoning, Jaynes explains, is a natural thought process. Logic, on the other hand, is "how we ought to think if objective truth is our goal." We need logic because reasoning is an unconscious process. Jaynes equates reasoning to logic as conduct is to morality. Our conduct can be an indicator of our underlying morality, just as our reasoning can be an indicator of our underlying logic.

Jaynes believes the rise of consciousness to be a result of the rise of language, and in turn the development of metaphors. The way in which we describe things affects how we believe they function. For example, the brain is often referred to as a computer, which affects how we believe the brain "connects", "processes", and "stores" information.

Jaynes outlines Six features of Consciousness that all tie into his core theory of metaphor and language.

1. Spatialization. Spatialization is the process in which we give objects "space" in our mind to exist. We separate concepts through the belief we are giving them space in our mind to exist.

2. Excerption. Excerption is the process of using a part to signify the whole. We allow a part of an object to represent the whole in our consciousness. Our perception of this object will affect what we excerpt. For instance, if we do not like a certain person, our excerpt of this person will most likely be a negative feature about that person.

3. The Analog "I". The Analog "I" is the metaphor we hold of ourselves in our consciousness.

4. The Metaphor "Me". The Metaphor "Me" is how we see ourselves when we take a step back from the situation in our consciousness.

5. Narratization. Narratization is the constant story of our lives that runs through in our mind when we are conscious. Narratization assists us in assigning causes to our behavior.

6. Conciliation. Conciliation is the process of making excerpts or narratizations compatible with one another. This is equitable to making internal conceptions and new stimulus agree with one another, or finding a way to categorize the new stimulus. For example, when asked to picture a meadow and a tower, we automatically tend to think of the tower in the meadow, as it is our only way to think of two things at once. However, when asked to think about a mountain and an ocean, we are only able to think of both at once through the process of narratization.

Hawkins' Hierarchical Temporal Framework

Hawkins proposes a neural model of the cortex which conceptually lies between the high levels of Jaynes' model and the lower level of individual neurons. He proposes that the lower level neurons organized into a hierarchical set of "nodes" (See Figure 1). Hawkins assumes "nodes" in the cortex perform four basic actions: Discover Causes, Inference, Prediction, and Behavior. Each "node" is probably a group of many neurons, but the specific map of Hawkins' model to individual biological neurons is not complete. We discover causes through learning from our external environment. If a loud noise arises when the door is closed, we discover the cause for the noise is the door closing. Once we have discovered causes, we can infer as to what will happen in the world. Now, if we see a door close, we can infer a loud noise will follow. Prediction is part of our feedback loop. We are often unaware that we are predicting actions. We only become aware of our predictions when they do not match what is happening in the world. For example, if a door closed and we heard a soft squeak instead of a slam, we would be surprised, not because of the squeak, but instead because of the lack of a slam. If the door slammed, we would not be phased and continue our current action. If the door squeaked, however, we would most likely look at the door and attempt to find the source of the squeak. Attention, consciousness, and prediction are all related in Hawkins' model.

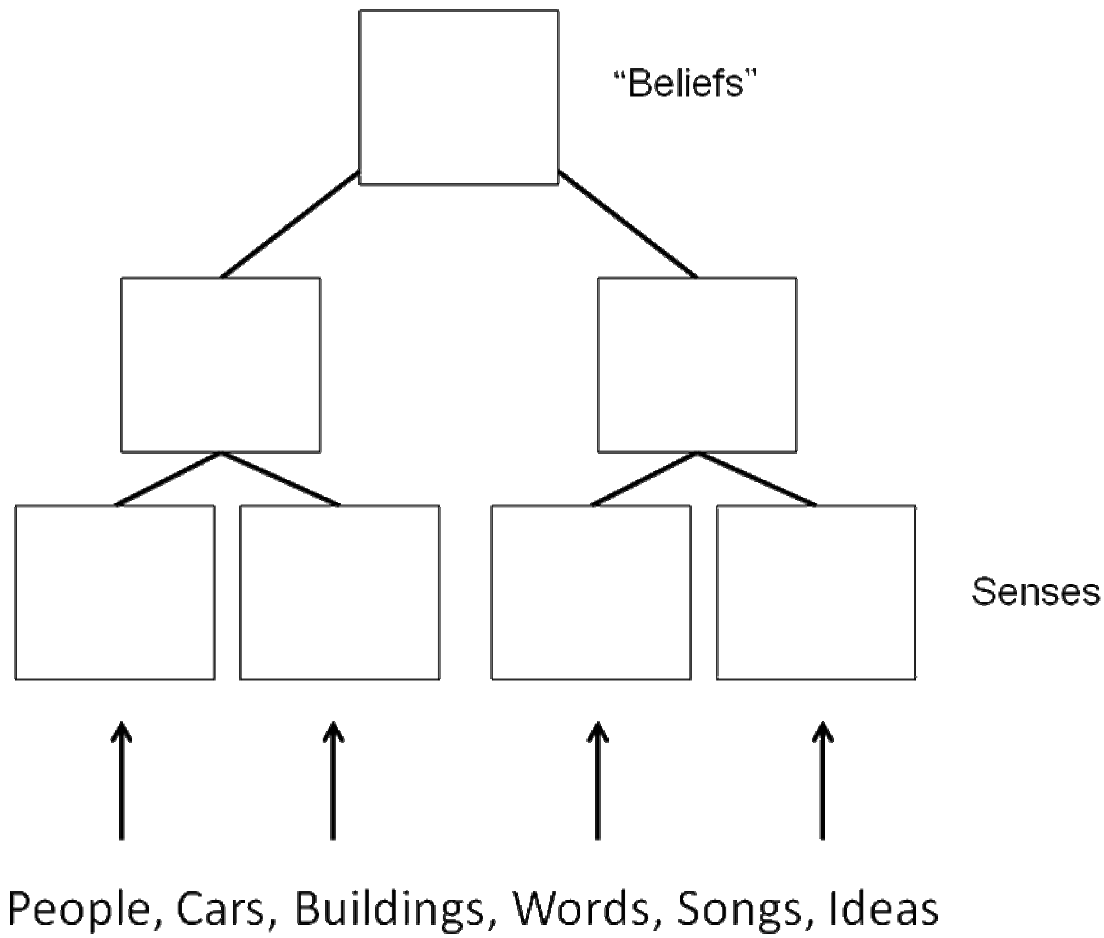

Figure One: This is a visual representation of Hawkins' Hierarchy. We receive input from the world, including people, cars, buildings, words, songs, and ideas, and we sense them at the lower level of the hierarchy. Once the senses are passed up the hierarchy as a set of patterns, we form beliefs about those patterns.

At the lower level of Hawkins' hierarchy, nodes only perceive a small part of the world, such as a line segment or a corner. Each node learns sequences of patterns and pass the sequence's name up the hierarchy. The lower level nodes feed information about their section to an upper level node, which puts together the input and compares it to previously learned patterns, thus a higher level node sees a larger piece of the world. When the node recognizes part of a pattern, it is fed upward in the hierarchy until the top node recognizes the entire pattern and action can therefore be taken. Each node performs exactly the same function (i.e. discovering causes, etc...), but differ in what input patterns they see. The lowest level nodes see a small part of the raw input from the world. The next level sees the output patterns from the first level, and thus sees a larger part of the world, processed through the first level. This continues up the hierarchy, each node discovering higher and higher-level causes and making predictions.

An example is how we perceive music. Music is a sequences of notes, with the lower level of the hierarchy seeing a few notes at a time, then a pattern develops from those notes, whether it be of four notes going up or down, and gives a name to the sequence. The nodes pass the name up, and combine it with other sequences, perhaps seeing an entire scale going up or down, and give a name to that sequence, and so forth until the hierarchy passes name of the song up.

There are three specific ways in which a node can learn a pattern and to recognize it. Firstly, a node can recognize a pattern from sequences of spatial patterns. Therefore, if we have a line progressively moving across the screen, our next logical guess is to assume the line will continue moving. The node then learns that pattern. Second, if a node sees a certain pattern over and over again, the node recognizes this might be something worth remembering. This is why repetition often helps people learn and remember little details of information. Thirdly, nodes use context from above in the hierarchy. Therefore, if what a lower level node is seeing is actually part of an overall pattern higher up in the hierarchy, the node is able to learn that part of the pattern and recognize it at a later point in time. Therefore, a node needs to use both time and space to learn. This is often seen when children pick things up and move it when there is a novel object.

The hierarchy allows for very focused attention on a specific object. It is impossible to focus precisely on more than one thing at once, as our mind has mechanisms for condensing the information together into one so we can think of both at the same time, which is very similar to Exerptation and Conciliation in Jaynes' Features of Consciousness. However, a person can stare at a certain part of an object, say a keyboard, focus on a specific letter, focus on a specific part of the letter, focus on a specific angle in the letter, and finally focus on a specific line. In another example, I can look at a painting with hundreds of people (think of Where's

Waldo), focus on a specific face, a specific facial feature, and finally focus on the brush stroke.

Finally, the hierarchical structure allows for belief propagation. Belief propagation is the process in which nodes on the bottom of the hierarchy quickly force the nodes higher up in the hierarchy to create a belief about the input. In just one pass, the hierarchy can create a belief about a pattern. Additionally, belief propagation resolves ambiguity among inputs quickly by looking for consistencies among input, and establishing the most consistent input as the most likely pattern or object.

Many of Jaynes' features of consciousness can be explained through Hawkins' hierarchy. Exerptation, for example, uses a part of the whole to signify the whole. This is the same as one node recognizing a part of a pattern, and knowing what it is through the hierarchy. Conciliation can force the lower level nodes to come up with an overall object very quickly, as it pulls sequences together.

<u>Neuroscience and Psychology</u>

The purpose of this paper is entrenched in the notion that psychology, neuroscience, and religion all affect one another, and their interconnectedness is key to understanding our emotions, actions, behaviors, feelings, and morality. Only in understanding the interconnectedness of psychology, neuroscience, and religion, can we effectively understand human nature. Necessary to this explanation is mirror neurons. Mirror neurons act as a link between individuals, exciting both when a person performs an action, and when a person watches someone else perform that action. This link could serve as an innate connection between humans.

Machamer and Sysma (2007) have a similar argument to my theory of connection between neuroscience and psychology. They work to defeat the notion that neuroscience cannot explain all of psychology, because it only looks at neural connections to behavior or cognitive events. They also propose neuroscience is relevant to psychology because of the inclusion of the historical importance of culture and human knowledge. Psychology is just as important to neuroscience as neuroscience is to psychology. Psychology provides neuroscience with an explanation for the cultural affect on our actions and behaviors. Neuroscience provides psychology with the "why" to our behaviors, not just the how and what. Genetics themselves evolve because of cultural phenomenon. Gene- Culture Coevolution theory explains: "biology… exerts pressures on the ways in which cultures develop, just as culture has exerted pressures on the evolution and development (in a life- time) of many of our neuro-mechanisms" (Machamer and Sysma 2007). Lactose is a fitting example. Because it is found in mother's milk, mammals have a high tolerance for lactose as babies, and that wears off as we become adults. Lactose is controlled by a single gene, and this gene is found in areas that have a long history of dairy farming. Therefore, people who live in areas with continued dairy farming are not as lactose intolerant as adults. They continue to argue that this evolution has played a part in human psychology. Evolution has a direct impact on social learning, in the form of mirror neurons. It is possible to assume mirror neurons developed from gene-culture coevolution for the development of communities, and the distinction between right and wrong developed for a self policing group, which in turn organizes hunting, gathering, protection, and population growth. This is very similar to Jaynes' theory of the development of knowledge and its connection to consciousness (Jaynes 1990).

Vedantam (2007) reports it is innate in humans to be "good". The primitive part of the brain that is activated by food or sex, the reticular formation, is also active when performing a "good" action. This could insinuate morality was an evolutionary "plus", something that assisted survival. If you covered someone else's back, they would cover yours, helping protect each other from the elements or greedy neighbors. The article also states that empathy is the foundation for morality. One cannot have morally sound actions without an understanding of the consequences to others.

David (2008) mentions mirror neurons in his simulation theory. With this theory, he states we use our own experiences to understand and predict other people's behavior. We look for patterns in other's behaviors and compare them to our patterns to establish the expected outcome. For example, we can better understand why someone pulls their hand off a stove quickly if we have burned our hand on a stove before. This is possibly equitable to having empathy for the pain they feel. Thus, we can say that mirror neurons are a basis for empathy, which implies that mirror neurons could be the foundation of morality (Vedantam, 2007). David also introduces the "sense of agency" as self consciousness, knowing you exist, what you are doing, and that you have control over your actions.

David (2008) also points out experiments have shown that morality does not just consist of decisions, but the processes by which they are reached. These experiments show that when part of the prefrontal cortex, which controls emotions, is damaged, patients determined the end justifies the means in a moral conflict. They did not consider the emotional meaning behind the decision, such as killing a child instead of man, or a woman instead of a man. David therefore concluded emotions are part of what we define as morality.

So, we must be conscious of ourselves to have a sense of agency, and we need a sense of agency to have empathy, and we need empathy for morality. Furthermore, we need morality to establish a sense of guilt, which is often remedied with organized religion. To create a tie to psychology, one of the popular archetypes is God. God is seen as our redemption for guilt, which is what we feel when our actions do not match our sense of what is morally good and right. The only way our actions would not match our moral judgment is through conflicting structures in the brain, as is pointed out by David (2008). When we have a hard time judging what is morally correct in a situation, it is because structures are competing with each other.

<u>Consciousness</u>

As stated before, consciousness is necessary for morality. Consciousness affects both how we approach morality, and how we talk about morality. Based on different theories of consciousness, we can develop different theories on the capacity and location of morality. The following researchers look to find the neural correlate of consciousness, which can effect where we believe morality is located, and in turn, which living creatures qualify as "conscious and moral" creatures.

Crick and Koch (2003) create a framework to find the neural correlate of consciousness. They define a framework as a guide or map explaining which brain structures actively work as a location for consciousness. They explain the cerebral cortex fills in and makes quick assumptions, creating a stereotyped response. This is analogous to pattern recognition and prediction. They listed 10 different brain activities that closely relate to Hawkins' Theory of Consciousness.

1. The homunculus, located in front of the brain, is not conscious at all, but instead picks up on the sensory stimulation of our imagination.

2. Zombie Modes of Consciousness are rapid stereotyped responses necessary for survival.

3. Neurons work in a coalition, some being excitatory and some being inhibitory. We are only conscious of what the winning coalition senses, which requires attention.

4. Explicit Representation states that a specific set of neurons are made to respond to a specific input, without further processing. Further processing is only necessary when our prediction of actions does not match the actual action observed. This is similar to Hawkins' belief propagation (Hawkins 2005).

5. Input travels up the visual hierarchy, where there might be a zombie mode, and then gets trickled back down after the higher levels have processed it. The height of the neural activity would depend on the amount of attention paid to the subject.

6. Driving and modulating connections in the coalitions of neurons. Driving has to do with the basal dendrites, and modulating has to do with back projections in the apical dendrites.

7. It is possible consciousness is only a series of snapshots. We perceive motion then as a series of snap shots, with motion created from the constant firing of neurons.

8. Our brain could use attention and the process of binding, or the collapsing of multiple parts of an object by the brain into one object. The thalamus is the organ of attention, and easily could be involved in this process. If multiple cortical neural networks do not overlap, multiple objects can be focused on at once because of binding.

9. Neurons need to do more than just fire at the same time in order to develop consciousness, the timing of firing is equally important.

10. The penumbra has certain neurons that respond to specific input, and uses the specialty in that area to create a "plan" for that input. For example, if the neuron "sees" a hammer, the neurons create plans for hammering.

Bodovitz (2008) defines consciousness as the continuity of experience. He suggests that consciousness exists in cycles because of the delay between sensory perception and conscious awareness. Without "putative conscious vectors", we wouldn't be able to tie cycles together, resulting in a still frame image of the world, analogous with visual motion blindness. He asserts that the most likely candidate for calculating these vectors is a part of the pre-frontal cortex (DLPFC). Changes in the DLPFC coincide with awareness of change. The mid DLPFC activity arises when subjects are required to monitor information within working memory, regardless of the nature of information (visual, auditory etc.). When we sleep, the DLPFC is inactive, and can be only slightly active when we are in the limbo between sleep and awake. Humans have a larger prefrontal cortex than other animals, which is hypothesized to account for differences between humans and animals. We have no consciousness or memory of being infants, and this could be because the DLPFC does not grow until adolescence. Therefore, there is no consciousness to remember because the cycles cannot be tied.

It is important to decipher between the source of consciousness and the "gateway to consciousness". It seems here that we have found a brain structure, the DLPFC, which assists in conscious thought and has a part in conscious thought and awareness, but does not provide the neural correlate for consciousness. The DLPFC does not create the "cycles of consciousness", it ties them together.

Because of these cycles, the delay, and the vectors, Bodovitz (2008) provides us with specific advantages and disadvantages. The advantages include: letting us focus on salient features (those that are changing), places conscious awareness apart and on top of the cognitive hierarchy, allowing for global feedback and coordination, and cancellation of "noise" by calculating the changes in net activity. The disadvantages to these cycles include: awareness of change after the fact and the lack of consciousness of conditions when they fade into a steady state.

It is well known that it is impossible to be conscious of multiple stimuli at once. (Have you ever tried seeing everything in a picture at once and remembering it all?) Gallace (2008) attempts to explain tactile consciousness while using visual consciousness as a reference point. In experiments, subjects were provided with multiple visual stimuli and multiple tactile stimuli. Gallace found that when there were only 1-4 stimuli in either field, they could be easily identified. However, when more stimuli were introduced, the accuracy diminished. Gallace inferred that subjects did not have a direct apprehension of the stimulus, but were able to create a pattern, such as a triangle or square, and therefore could infer 4 points. Therefore, subjects grouped stimulus into 1 pattern. When similar information was entered through tactile and visual sensory, response may be enhanced. This explains why people usually can't look when being poked by a needle, or children make a game of guessing what letter is being drawn on their back. Because we can't see the tactile stimulus, its effect is diminished. This can be related to the interaction displayed in Hawkins hierarchy between audio, visual, and tactile, when all overlap and work together at the top to form the central pattern.

Gallace (2008) also mentions the possibility of stream of consciousness versus the cycle idea presented by Bodovitz. Gallace asserts that events separated by 20-50 ms are judged to be simultaneous when presented visually. The same sort of delay was found in tactile stimulus. Furthermore, Gallace summarizes Bliss's research which suggests the existence of super short term memory for tactile and visual events, which might sustain our consciousness. The short term memory quickly disintegrates.

Lau (2008) explains that just because a person can identify an object does not mean they consciously perceived it. In forced-choice detection or discrimination tasks, consciousness was not necessary. This was proved through patients who have blindsight. Blindsight is a phenomenon when patients who are blind in certain parts of their visual field can still perceive objects in the blind spots of their visual field (Carey 2008). Consciously, these patients cannot see what was in front of them, but when forced to chose between two stimuli, they chose correctly well about guessing percentage. Lau also talks about Signal Detection Theory. Signal Detection Theory states that there is a threshold for conscious recognition of an object. However, people have an idea as to their relative threshold, and if they are wrong on their assumption, researchers receive false positives or negatives. Therefore, this threshold setting could affect our perceptual consciousness.

Learning is important in setting the threshold or criteria for consciousness. We learn over time where to place our threshold so we can achieve the most correct answer most often. When we recognize a certain type of light, we have a response for it. If that response is wrong, we try to learn a new response. Lau notes ambiguity, dynamic fluctuation, and misrepresentations as ways we can incorrectly learn a response and therefore the criteria (Lau 2008).

<u>Mirror Neurons</u>

Mirror neurons are neurons found in different parts of the brain that activate both while watching an action being performed, and performing the action. Therefore, when I open a door, the neurons that excite in my brain excite in the person watching the door being opened. Mirror neurons play an important role in my hypothesis because they are a physical connection among all humans and monkeys, and they allow a link between the action of others and my actions. This is a prerequisite for morality.

Fadiga et al. (1995) conducted an experiment to identify if the same process of stimulation was involved in the human brain as it is in monkeys. “Area F5” in monkeys has neurons that fire during goal directed motors acts like grasping, manipulating, holding, and tearing. They used transcranial magnetic stimulation (TMS) to monitor brain activity and stimulation during observation of different motor acts. TMS excites neurons with weak electrical currents created by rapidly changing magnetic fields. Essentially, it painlessly activates one part of the brain so the circuitry and connectivity of the brain can be monitored. They found that the excitability of the motor system increases when a subject is watching a motor action. Also, due to the set up of the experiment, they could conclude that this effect depended on observation, not the possibility of having to repeat the action. Most importantly, the observation of an action stimulates neurons that would normally be used when physically performing the action, meaning that motor systems aren’t only devoted to physically doing the action, but also in the recognition of the movement. Because of the recognition of movement, we might be able to "see" ourselves doing the motion because our brain is using the same neurons. This is part of our prediction framework. Using the same neurons that we would to physically open the door, we can think about what we would do to open the door, predict what we would do. When the doorknob is moved, we are surprised because our movement and prediction do not match what is actually happening. This is related to Narratization in Jaynes' features of consciousness. When the story in our mind doesn't match the story of the world, we are startled into recognizing the difference and reformulating the story in our mind.

Buccino (2001) restates that mirror cells are located in the premotor cortex in Area F5 in monkeys, and further proved the existence of mirror neurons in humans. However, Buccino wanted to ensure that the part of the brain that controlled foot movement would excite when watching foot movement, and the same with mouth movement. Ultimately, they discovered that, indeed, regions activate in a somatotopic manner. This means that these mirror neurons are not solely discovered in one area of the brain, and don't represent just one action, but represents a plethora of observations. These observations could strengthen certain connection weights, and weaken others, which could explain why something can be considered the norm in one society, but immoral in others.

McGeoch (2007) uses the disorder Apraxia to display the brains ability through mirror neurons and metaphor. Apraxia is a disorder that makes it impossible to physically recreate the action shown in front of you. The person can explain their intentions, and what they saw, however, they cannot physically do it. The Inferior Parietal Lobule (IPL) is very important to the suggestion of mirror neurons and where they exist. McGeoch makes clear that mirror neurons are only activated when an action happens, not a single act. So just grasping will not activate the neurons, but grasping and putting the cup to the mouth and drinking activates the neurons.

McGeoch (2007) hypothesizes that mirror neurons within the human supramarginal gyrus have interconnected patterns that encode learned, skilled action. This is because when this area is damaged in apraxia patients, they cannot perform the action or can even tell if someone else is performing the action correctly. Patients don't have any prediction to base feedback on. It also is imperative for metaphor development, which McGeoch calls abstract types of re-conceptualization. Metaphors are a way to re-conceptualize patterns we already have, and apply the underlying ideas and conceptions to other things, creating a metaphor.

QUESTIONNAIRE

I wanted to test whether different religions responded differently to moral situations, and the extent of intergroup differences. Important to the purpose of this paper is defining the difference between right and wrong. Many believe their moral system of beliefs come from their religion, and for that reason, I wanted to measure any differences among religions. I thought it would interesting to see if Catholics believe the death penalty is moral while an Agnostic believes it is just as cruel as killing someone on the street in cold blood. A full copy of the questionnaire is provided in the appendix. Since the context in which we learn patterns matter, I provided the same action in different contexts to determine if there was a significant difference in response. I hypothesize that there will be significant differences in responses between contexts, yet the responses will be similar between religions per item.

The questionnaire consisted of a self report section, for classification purposes, and a statement rankings section. Firstly, the subject was presented with this statement:

> This questionnaire is informal in nature. You will not be judged by your responses. However, your responses will assist me in determining public opinion on morality. Please answer truthfully. I ask for age, gender, and religious belief system for classification purposes.

Afterwards, participants were asked to type in their age, and select their gender. There were 100 participates, with a total of 90 completing the survey. The average age of respondents was 34.5, with a median age of 22. 52.1% of respondents were male with 47.9% being female. Participants were also asked to identify their belief system, in an attempt to distinguish religion's effect on determination of moral and immoral actions, and the influence it might have on the understanding of right versus wrong. Catholic was the most frequent religion, with 42.4%, followed by Protestant, Agnostic, and Jewish with 23.9%, 17.4% and 8.7% respectively. Let me note, however, that this study was confined mostly to Bryant Faculty, staff, and students, with few exceptions. Due to this, the sample of religions does not accurately represent the percentages in the general population of the United States.

To determine where certain religions establish boundaries between right and wrong, I asked questions that were similar without context, but with context resulted in different situations. For example, I asked if killing people in other countries, killing Americans, and the death penalty were right or wrong actions. Stripped of context, all these actions are the same, as someone is killed as the result. However, are certain religions primed to believe one is right, while the other is wrong? To test this, I assigned number to the wrong-right scale (with one being completely wrong up to five being completely right), and separated by religion to see trends.

Because of a smaller sample size, I chose to use a statistical analysis to test the difference between my results, and the results that would have come about from a random sample (Jaynes, 2003). The Jaynes Psi function is as follows:

$$\psi_B = 10\sum_{k=1}^{m} n_k \log_{10}\left(\frac{n_k}{np_k}\right)$$

where:

- N_k= number of people who responded within the category
- N= total number of responses
- P_k= the probability you would have expected if it was a random sample

The larger the result, the more different the results of the survey is from a random result. Significance was defined at a 5% level. Respondents were asked to respond on a 5 point scale, 1 being completely wrong, and 5 being completely right. The numbers were clearly labeled in each question. The results are as follows:

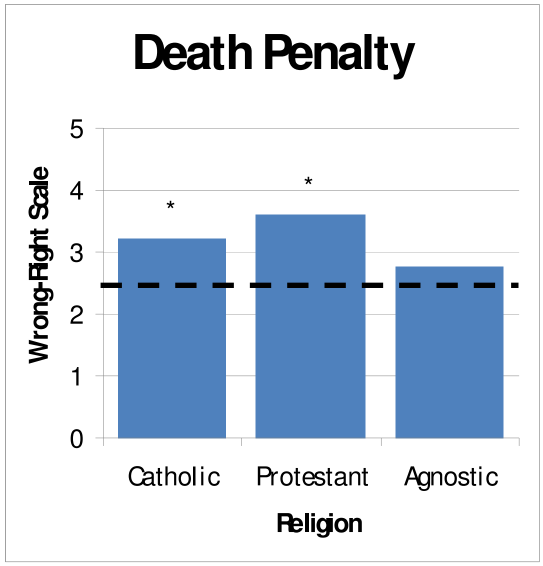

*Figure Two: Significance was set at 5%, denoted by a single *, 1% is denoted by a double **. The dashed line represtns 2.5, or the middle of the scale. These facts hold true for all figures.*

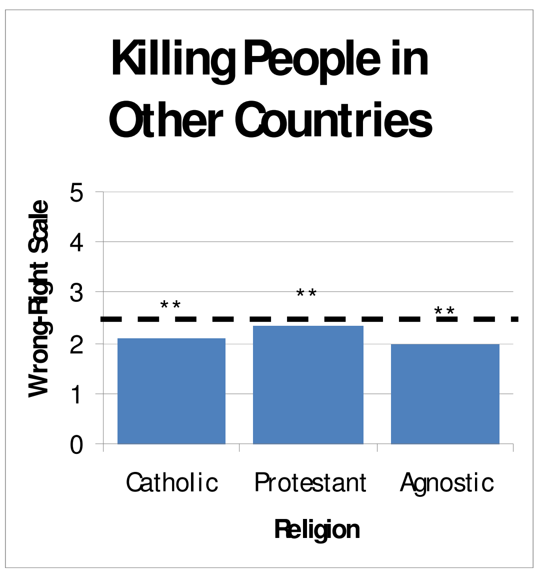

Figure Three

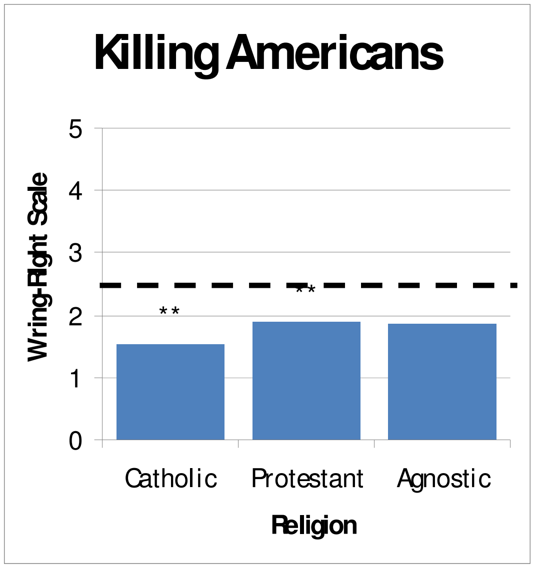

Figure Four

As we can see, the top three religions, in general, overwhelmingly agree for the purposes of this analysis. However, it is clear that Protestants do not see killing others, no matter the context, to the same severity Catholics and Agnostics do. This could be for various reasons. For example, because Catholics distinguish between mortal and venial sins, killing someone else is an unforgivable sin that condemns you to hell, while Protestants do not make that distinction among sin. Also, Agnostics tend to follow the Golden Rule, "Do unto others as you wish to have done to you." This saying, along with the results, insinuates Agnostics see a greater severity in harming others. However, there was only statistical significance in "Killing People in Other Countries" for Agnostics, so it is not fair to comment on the other results.

The other interesting questions inquired about stealing. Respondents were asked if they would classify stealing food for a pet, stealing food for a dog, or stealing electronics from a store were wrong or right. When stripped of context, all these actions include stealing. Stealing is part of the Ten Commandments, "Thou Shall Not Steal." I did not graph the results for stealing electronics because 100% of respondents said the action was somewhat wrong or completely wrong. The results are as follows:

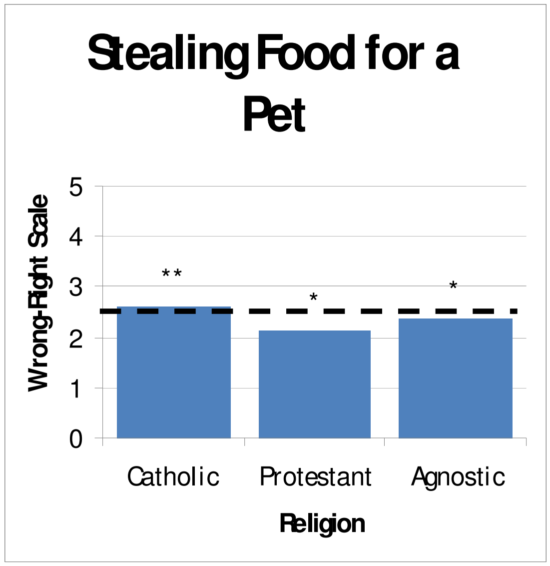

Figure Five

Figure Six

Here we can see "Stealing for your pet" and "Stealing food for your family" are very similar in nature. However, once again, Protestant and Agnostic respondents to "Stealing Food for Your Family" was not of statistical importance.

Overall, as predicted, there were differences among context, but little difference among religions. However, we did see a different pattern of response among religions. These results speak to the influence religion has over perception of wrong and right, and the "hold" a religion has over how strong one feels about a certain subject. For instance, Catholicism, on the basis of these results, seems to have overwhelming control over patron's beliefs of wrong and right, where Protestantism has different beliefs, or a belief system that encourages free thought within a set of guidelines. Agnostics, who do not identify with a religion yet maintain a set of moral beliefs seem to uphold their own moral judgment of the golden rule "Do unto others as you will have done to you." This results shows a fundamental similarity between religion's perspectives, and allows me to continue using a presumed Judeo-Christian perspective in defining morality, while leaving room for slight variation.

To relate this to Jaynes' theory of metaphors, therefore, the language used to define right and wrong in the context of different religions changes how individuals conceptualize right and wrong. For example, in the Catholic church, there are moral sins, and there are venial sins. Therefore, killing someone is a different level of wrong than lying to your mother. Protestants, on the other hands, classify all sins under one umbrella. This language distinction can affect how sever different religions classify criminal actions. Hawkins theory, on the other hand, states context and recognition are important, which would assume the context in which these people experience these situations affected their answer.

This also shows us a strong connection between mirror neurons and culture. The formation of patterns and recognition only happen within the context of our lives. Although mirror nuerons are a biological factor, this biological factor only fits within the context of culture.

COMPUTATIONAL MODEL

For the purposes of this paper, we have used an associative network as a basis for our mirror neuron learning process (Bechtel, 2002). The associative network represents different clusters of specific attributes of a subject. Here is an example:

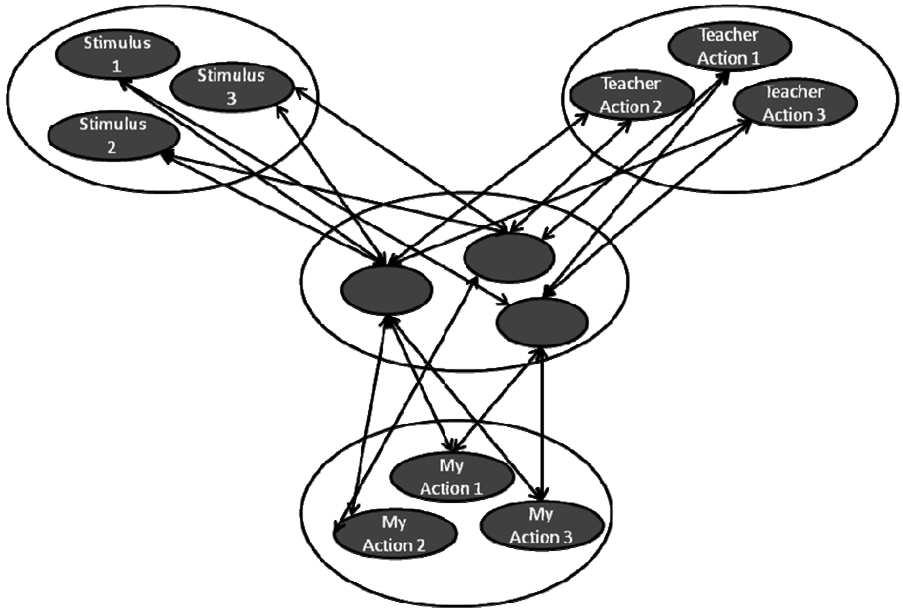

Figure Seven: This associative network displays the possible random connections between Stimuli and Actions, connected through a hidden layer of neurons.

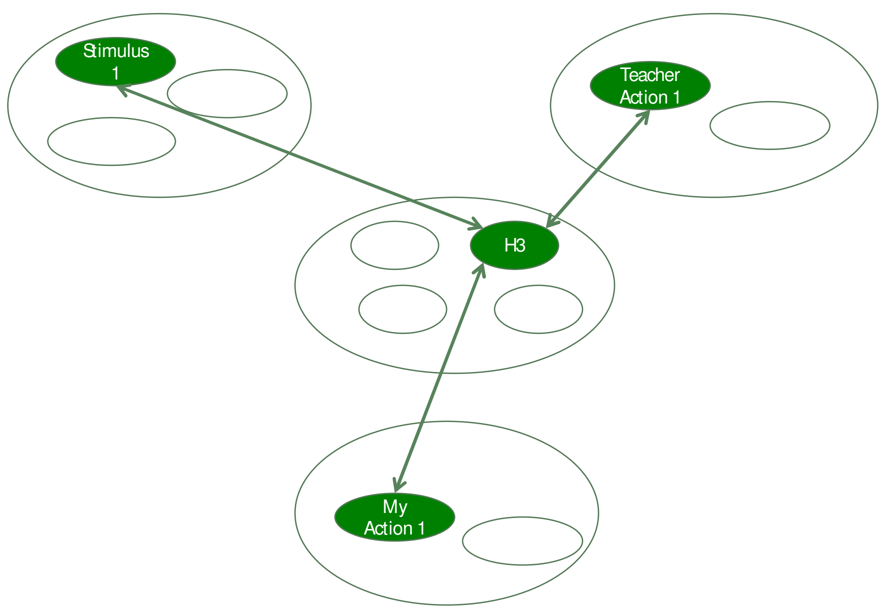

Figure Eight: Here we can see the associative network learned the connection of Stimulus 1, Teacher Action 1, and My Action 1 thought Hidden Unit 3, which now is a mirror neuron.

For example, we have a cluster that is the stimulus (what is going on in the world), a cluster that represents the teacher, and a cluster that represents my own personal actions. When a teacher performs an action, and it is in a certain context (world state/stimulus), and my actions represent those of the teacher, the hidden units create a stronger connection between the units, increasing the weights between units. After many repetitions of this "learning", the weights between attributes and the hidden layer become stronger, creating a prediction of attributes if only one attribute was presented.

To equate this example to mirror neurons, mirror neurons are created when the hidden neurons in the associative network learn the correct connections and weights. For the purposes of this model, we started with random weights and connections among clusters. We have a cluster that is the stimulus (what is going on in the world), a cluster that represents the teacher, and a cluster that represents my own personal actions. When a teacher performs an action, and it is in a certain context (world state/stimulus), and my actions represent those of the teacher, the mirror neurons create a stronger connection between the units, increasing the weights between units. When one of those units in not activated, but two of them are, the mirror neurons, in turn, activate the third, resulting in recognition. Without learning, the neurons wouldn't be able to do that. This could be a recognition of an action, or it can be a recognition of a feeling (pain, happiness, angry, frustration). This recognition is empathy. This is only one argument as to the formation of mirror neurons, however. It is also fair to argue mirror neurons arise from an evolutionary survival mechanism. Those who stick together have added protection, better ability to catch prey, and the ability to reproduce, building a civilization.

<u>Method</u>

We started the simulation to see if we could obtain mirror-neuron like behavior at all to arise from a simple environment and learning paradigm. The paradigm was based off Bechtel's (2002) equations for associative networks. Net input to each unit had to be calculated for this model. Net input is all the input to a specific unit. Special parameters must be used in this example, because we cannot simply add the connections and multiply them by their weights because the unit is in connection with the external environment. To calculate the net input, we used the following equation, in which .1 and .4 are given default values:

$$netinput_u = 0.1\sum_i weight_{ui} output_i + 0.4 extinput_u$$

where:

- u = the unit whose activation is being calculated
- i = other units that are feeding into u

The next important equation is the change in activation of a unit. This equation serves as a simple activation rule. The equation is as follows:

$$\Delta a_u = (max - a_u)\ (netinput_u) - (decrayrate)\ (a_u - rest)$$

where:

- Δa_u = the net current change to be made to $activation_u$
- max= maximum activation value that a unit can take

There are two parts to this equation. The first part calculates the change from the net input, while the second serves as a decay variable that decreases activation, even where there is no net input.

At each iteration, we activate a single stimulus, my action, and teacher action simultaneously. We allow the network to converge to its equilibrium, and then adjust the weights of the network in parallel. Because the stimulus, my action, and teacher action are always present in Trial 0, it is not too surprising that the network learns to associate them. The connections and probabilities for Trial 0 were as follows:

	My Action	**Teach Action**	**Stimulus**
	1	1	1
	2	2	2
	3	3	3
Probability	100%	100%	100%

Table 1- Trial 0

Trial 1 tested the necessary existence of variables to represent learning. With a one to one relationship among stimulus and actions (i.e. stimulus to action pairs of 1 to 1, 2 to 2, and 3 to 3), My Action, Stimulus, and Teacher Action were set to different probabilities of representation. This was in order to test the level at which learning can still happen. Each probability was separately set in succession to a point where learning no longer occurred. The connections and probabilities for Trial 1 were as follows:

	My Action	**Teacher Action**	**Stimulus**
	1	1	1
	2	2	2
	3	3	3
Probability	75%	84%	87%

Table 2- Trial 1

Trial 2 tested if the program could learn with multiple stimuli to one action (i.e. 1 *and* 2 to 1, 3 to 2 and 4 to 3). This represents a more realistic picture of the world, as we are constantly combated with multiple stimuli at once, and are forced to weed out the noise and find the real pattern. Once the program proved an ability to learn with multiple stimuli, the probability of the actions and stimulus being present were reduced to establish a point in which there was no longer learning. The connections and probabilities for Trial 2 were as follows:

	My Action	Teacher Action	Stimulus
	1	1	1 ***and*** 2
	2	2	3
	3	3	4
Probability	10%	10%	15%

Table 3- Trial 2

Each trial results in a graph color coding the strength of connection between actions and stimuli per hidden neuron. This graph not only shows which actions and stimuli are connected, but also shows which hidden neurons connect them. The trial was also programmed to run for multiple iterations per trial, representing the time and repetition necessary to learn patterns. The weights in the graph changed over time until they settle, representative of learning. The lower left hand corner and upper right hand corner represent the learning. The darker red the box is, the heavier the weight between stimuli, action, and hidden unit is. The blue represents an inhibitory (negative) connection.

An example of this graph is as follows:

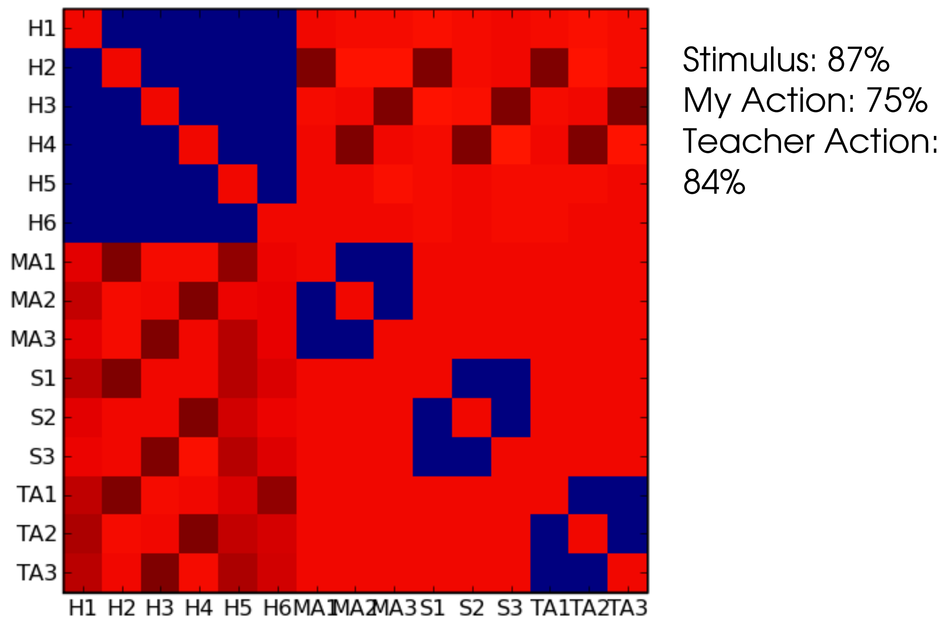

Figure Nine: Here we can see a representation of a strong connection, a weak connection, and an inhibitory connection. For example, My Action 1 is connected to Stimulus 1 and Teacher Action 1 through Hidden Unit 2.

Here we can see that My Action 1 (MA1), Stimulus 1 (S1), and Teacher Action 1 (TA1) connect through Hidden Unit 2 (H2). This means the computer has learned the connection among these three variables, and in the future could recall the pattern if only given part of the pattern.

When the program did not sufficiently learn the pattern, it resulted in a pattern as follows:

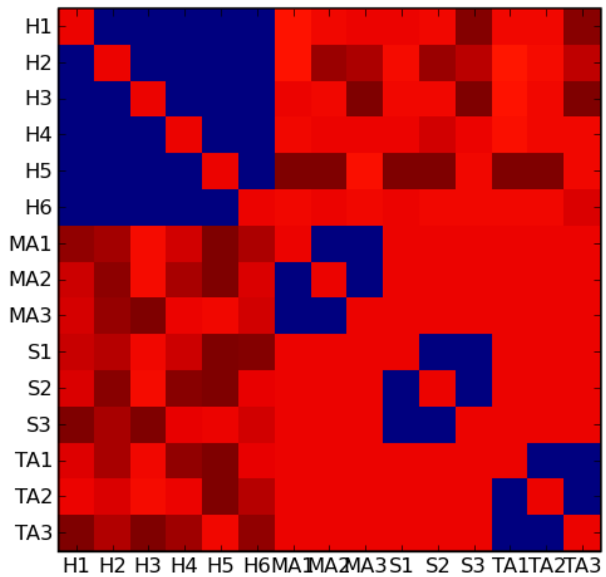

Figure Ten: Here we can see there was no convergence among units. The weights are still noisy and connection are incorrect.

<u>Results</u>

In the first batch of trials, it was found that the program was successful in learning the designated pattern as long as the stimulus and actions were represented together at least half the time. That is to say, as long as the probability of representation of My Action, Stimulus, and Teacher Action multiply to be greater than about .5, the pattern will be learned (See Table 2). For example:

```
units[stim_name].external_input=1.0 if rand()<.87 else 0
units[my_name].external_input=1.0 if rand()<.75 else 0
units[teach_name].external_input=1.0 if rand()<.84 else 0
```

results in an overall probability of all three units being activated at the same time of 55%. However, if the probability is 40% or lower, the pattern is generally not learned. This 50% threshold is not a hard threshold. This can change to 48%, 51%, or any other type of variation within reason. When the program fails to learn, multiple variables become paired in an incorrect manner.

The second set of trials was conducted testing multiple stimuli to one specific action. In this example, Stimulus 1 and 4 were connected to Actions 1, while the rest were paired in a one to one relationship (i.e. 3 to 2, 4 to 3)(See Table 3). An interesting phenomenon arose through these trials. Instead of having a threshold, similar to the first batch of trials, there was nearly no floor to the probabilities of representation. When just one variable was changed (stimulus), the probability could be lowered all the way to .01 before the program failed to learn the pattern. When all three variables had changed probabilities, the probabilities were lowered to the following before the computer program failed to learn the pattern:

```
units[stim_name].external_input=1.0 if rand()<.15 else 0
units[my_name].external_input=1.0 if rand()<.1 else 0
units[teach_name].external_input=1.0 if rand()<.1 else 0
```

This means that the pattern was shown simultaneously only.15% of the time, and the program still managed to learn the pattern over time.

<u>Discussion</u>

When there was a simple one to one relationship among stimulus and actions, a threshold of about 50% in order for the program to learn and recognize the pattern. However, when multiple stimuli were present, which is more representative of the real world, the threshold arose around .15%. This implies the flexibility of our brain to learn patterns over time while different "white noise" is present at the time of learning.

The ability of the program to learn patterns with multiple stimuli without consistent representation provides insight to the flexibility and power of the brain. We, as humans, are able to learn a new pattern in the world, over time, when all the actions and stimuli are not present at the same time. Time, in this context, simply represents repetition. The more times one is presented with the pattern of some kind, the stronger a connection to learn the pattern. This could help explain why some are better at reading facial expressions than others. If one has had more experience with facial expression and the underlying emotions associated with them, the more likely one is to learn that pattern.

This also speaks to the opportunities one has to learn certain patterns in their lifetime. Culturally, it is more acceptable for a girl to cry than a boy. Girls are given more opportunities to understand emotion and how to read hurt, anger, frustration, happiness, and loneliness in other people. This pattern recognition is empathy. Often, women are said to be more empathetic than men. This could be because men are culturally taught to ignore emotion and suck it up, while women are taught it is "ok" to cry, giving women more opportunities in life to learn the pattern of external stimulus (emotion), my action (sympathy) and the teacher action (facial expression of other).

CONCLUSION

Mirror neurons, through their use of pattern recognition and repetition, work as a fundamental connection between individuals. This connection and understanding of feeling, and in turn sympathy, is defined as empathy. Empathy can also be defined as the foundation of morality. When an individual is able to understand the emotional and physical consequences of an action on another, their actions are said to be moral in nature.

Through repetitive exposure to patterns in life, we learn empathy. Some, depending on culture gender, and other demographic factors, have different exposure to empathetic patterns. The computational model provides us with a real world representation of pattern learning and recognition over time. The program was able to learn when multiple stimuli were paired with one action, and the stimulus and action representation were low. This scenario is more similar to the world, as there are multiple stimuli at once while we learn patterns.

Morality and consciousness can only be expanded to other animals depending on the theory one applies. If one views consciousness and morality through Jaynes' perspective, language is necessary for consciousness, limiting consciousness and morality to humans. However, if one views consciousness and morality through Hawkins' perspective, pattern recognition is necessary for consciousness, expanding consciousness and morality to all animals with a cortex.

This paper attempted to draw a tie between psychology, computational neuroscience, religion and philosophy. Through an extensive literature review, social psychology theory, a questionnaire regarding differences in judgments along religious lines, and a computation model, I have tried to show that only through the understanding and incorporation of all these aspects can our final goal be achieved.

There are some inherent limitations to this paper. First, I only studied a low level of neurons. The testing I conducted was not incredibly sophisticated in terms of direct applicability to creating an empathetic robot. Also, I was unable to directly test human responses with an fMRI machine. The assumption of mirror neurons in humans only comes from my literature review, and I was unable to prove it through my own experimentation. Finally, how the brain specifically operates is still unknown. Although we can identify some processes, the multiple connections and intricacies of the brain are still a mystery.

Future studies could be conducted using an fMRI machine to test the location of mirror neurons in the brain, and what types of stimulus is necessary to activate them. Also, more advanced program could be used, that could “read” an actual visual pattern paired with an audio input, to represent real stimuli. Also, one could test the differences among demographics in empathetic response, and the brain activity involved in making empathetic decisions.

APPENDICES

Appendix A – Questionnaire

This questionnaire is informal in nature. You will not be judged by your responses. However, your responses will assist me in determining public opinion on morality. Please answer truthfully. I ask for age, gender, and religious belief system for classification purposes.

1. Age:

2. Gender

- Male
- Female

3. **How do you identify your belief system?**

- Protestant
- Catholic
- Jewish
- Muslim
- Buddhist
- Hindu
- Atheist
- Agnostic

Other (please specify)

1. Please categorize the following statements

I am a moral person	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
I know the difference between right and wrong	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
I am more moral than my friends	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
Morality arises from biological processes	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
My morality is a result of my religion	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
My morality	Completely	Somewhat	Neutral	Somewhat	Completely Agree

is a result of my family	Disagree	Disagree		Agree	
My morality is a result of my school	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
My morality is a result of my friends	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
My morality is a result of myself	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree
Immoral people are able to be rehabilitated to be moral	Completely Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Completely Agree

2. Please rate the following situations as right or wrong

The Death Penalty	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Killing people in other countries	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Killing Americans	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Stealing food for a pet	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Stealing food for your family	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Stealing electronics from a store	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Telling a lie	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Using previous class tests to prepare for a current test	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Seat hopping at a concert	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right
Downloading music for free	Completely Wrong	Somewhat Wrong	Neutral	Somewhat Right	Completely Right

Appendix B- Gang Network.py Code

```
from pylab import *

from matplotlib import rc
size=20
family='sans-serif'

rc('font',size=size,family=family)
rc('axes',labelsize=size)
rc('axes',titlesize=size)
rc('xtick',labelsize=size)
rc('ytick',labelsize=size)
rc('legend',fontsize=size)


class Unit(object):

    def __init__(self,label):

        self.resting_activation=-0.1
        self.label=label
        self.activation=self.resting_activation
        self.external_input=0.0
        self.maximum_activation=1.0
        self.minimum_activation=-0.2
        self.decay_rate=0.1
        self.internal_strength=0.1
        self.external_strength=0.4

        self.netinput=0.0

        self.learning_rate=0.0

        self.initial_w=[]

        self.w=[]
        self.i=[]

        self.activation_list=[self.activation]

    def output(self):

        if self.activation>0:
            return self.activation
        else:
            return 0.0


    def reset(self,keep_weights=False):

        self.activation=self.resting_activation
        self.external_input=0.0
        self.netinput=0.0
        self.activation_list=[self.activation]
        if not keep_weights:
```

```
            self.w=self.initial_w[:]

    def update_netinput(self):
        self.netinput=0.0
        for (i,w) in zip(self.i,self.w):
            self.netinput=self.netinput+self.internal_strength*w*i.output()


self.netinput=self.netinput+self.external_strength*self.external_input

    def update_activation(self):

        if self.netinput>0:
            self.activation=self.activation+(
              (self.maximum_activation-self.activation)*self.netinput -
              self.decay_rate*(self.activation-self.resting_activation))
        else:
            self.activation=self.activation+(
              (self.activation-self.minimum_activation)*self.netinput -
              self.decay_rate*(self.activation-self.resting_activation))


        self.activation_list.append(self.activation)

    def update_weights(self):

        for u in range(len(self.i)):
            if self.w[u]>=0.0:  # only learn excitatory

self.w[u]+=self.learning_rate*self.netinput*self.i[u].output()
                if self.w[u]>1.0:
                    self.w[u]=1.0

                if self.w[u]<0.01:
                    self.w[u]=0.01

                # make symmetric
                j=self.i[u].i.index(self)
                self.i[u].w[j]=self.w[u]

        # sum_square=sqrt(sum([w*w for w in self.w if w>=0.0]))
        # for u in range(len(self.i)):
        #     if self.w[u]>=0.0:  # only learn excitatory
        #         self.w[u]/=sum_square



    def plot_activation(self):

        t=range(len(self.activation_list))
        plot(t,self.activation_list,linewidth=3)


    def details(self):
        print "My Label: ",self.label
```

```
        print "My Current Activation,",self.activation
        print "Connected to:"

        conn=zip(self.w,self.i)
        conn.sort(reverse=True)

        for w,i in conn:
            print "  ",i.label,':\t w=',w


    def __repr__(self):

        s="\tLabel: %s, Activation: %f\n" % (self.label,self.activation)

        s=s+"\t\tConnected to: "
        for (i,w) in zip(self.i,self.w):
            if w>0.5:
                s=s+i.label+" "
        s=s+"\n"

        weakly=False
        for (i,w) in zip(self.i,self.w):
            if 0.0<w<=0.5:
                weakly=True
                break
        if weakly:
            s=s+"\t\tWeakly Connected to: "
            for (i,w) in zip(self.i,self.w):
                if 0.0<w<=0.5:
                    s=s+i.label+" "
            s=s+"\n"

        s=s+"\t\tInhibiting: "
        for (i,w) in zip(self.i,self.w):
            if w<0:
                s=s+i.label+" "
        s=s+"\n"

        return s

def connect(unit1,unit2,weight=1):

    unit1.i.append(unit2)
    unit1.w.append(weight)
    unit1.initial_w.append(weight)

    unit2.i.append(unit1)
    unit2.w.append(weight)
    unit2.initial_w.append(weight)

def disconnect(unit1,unit2,weight=0):

    ret=False
    for i in range(len(unit1.i)):
        if unit1.i[i]==unit2:
```

```
            unit1.w[i]=0
            unit1.initial_w[i]=0
            ret=True

    for i in range(len(unit2.i)):
        if unit2.i[i]==unit1:
            unit2.w[i]=0
            unit2.initial_w[i]=0
            ret=True


    return ret


def connect_cluster(units,labels):

    for i1 in range(len(labels)):
        for i2 in range(i1+1,len(labels)):
            u1=units[labels[i1]]
            u2=units[labels[i2]]

            connect(u1,u2,-1)

def get_randomly_connected_units():


    gang_info=get_gang_info()

    names=gang_info.keys()
    names.sort()
    members=[s.upper() for s in names]

    education=[]
    for p in gang_info:
        if gang_info[p]['education'] not in education:
            education.append(gang_info[p]['education'])
    education.sort()

    age=[]
    for p in gang_info:
        if gang_info[p]['age'] not in age:
            age.append(gang_info[p]['age'])
    age.sort()

    occupation=[]
    for p in gang_info:
        if gang_info[p]['occupation'] not in occupation:
            occupation.append(gang_info[p]['occupation'])
    occupation.sort()

    marital_status=[]
    for p in gang_info:
        if gang_info[p]['marital status'] not in marital_status:
            marital_status.append(gang_info[p]['marital status'])
    marital_status.sort()
```

```
    gang=[]
    for p in gang_info:
        if gang_info[p]['gang'] not in gang:
            gang.append(gang_info[p]['gang'])
    gang.sort()

    #=============== make the network

    units={}

    for cluster in
[names,members,education,occupation,marital_status,gang,age]:
        for label in cluster:
            units[label]=Unit(label)

        if not cluster is names:
            connect_cluster(units,cluster)

    for cluster in [names,education,occupation,marital_status,gang,age]:
        for label in cluster:
            for member in members:
                connect(units[label],units[member],rand()*0.1)  # connect
to all

    return units


def update(units):
    for label in units:
        units[label].update_netinput()
    for label in units:
        units[label].update_activation()

def update_weights(units):
    for label in units:
        units[label].update_weights()

def reset(units,keep_weights=False):
    for label in units:
        units[label].reset(keep_weights)

def reset_external_input(units):
    for label in units:
        units[label].external_input=0.0

def get_gang_info():

    gang_info={}

    gang_info['Art']={}
    gang_info['Art']['gang']='Jets'
    gang_info['Art']['age']='40s'
    gang_info['Art']['education']='J.H.'
    gang_info['Art']['marital status']='Single'
```

```
gang_info['Art']['occupation']='Pusher'

gang_info['Al']={}
gang_info['Al']['gang']='Jets'
gang_info['Al']['age']='30s'
gang_info['Al']['education']='J.H.'
gang_info['Al']['marital status']='Married'
gang_info['Al']['occupation']='Burglar'

gang_info['Sam']={}
gang_info['Sam']['gang']='Jets'
gang_info['Sam']['age']='20s'
gang_info['Sam']['education']='COL.'
gang_info['Sam']['marital status']='Single'
gang_info['Sam']['occupation']='Bookie'

gang_info['Clyde']={}
gang_info['Clyde']['gang']='Jets'
gang_info['Clyde']['age']='40s'
gang_info['Clyde']['education']='J.H.'
gang_info['Clyde']['marital status']='Single'
gang_info['Clyde']['occupation']='Bookie'

gang_info['Mike']={}
gang_info['Mike']['gang']='Jets'
gang_info['Mike']['age']='30s'
gang_info['Mike']['education']='J.H.'
gang_info['Mike']['marital status']='Single'
gang_info['Mike']['occupation']='Bookie'

gang_info['Jim']={}
gang_info['Jim']['gang']='Jets'
gang_info['Jim']['age']='20s'
gang_info['Jim']['education']='J.H.'
gang_info['Jim']['marital status']='Divorced'
gang_info['Jim']['occupation']='Burglar'

gang_info['Greg']={}
gang_info['Greg']['gang']='Jets'
gang_info['Greg']['age']='20s'
gang_info['Greg']['education']='H.S.'
gang_info['Greg']['marital status']='Married'
gang_info['Greg']['occupation']='Pusher'

gang_info['John']={}
gang_info['John']['gang']='Jets'
gang_info['John']['age']='20s'
gang_info['John']['education']='J.H.'
gang_info['John']['marital status']='Married'
gang_info['John']['occupation']='Burglar'

gang_info['Doug']={}
gang_info['Doug']['gang']='Jets'
gang_info['Doug']['age']='30s'
gang_info['Doug']['education']='H.S.'
```

```
gang_info['Doug']['marital status']='Single'
gang_info['Doug']['occupation']='Bookie'

gang_info['Lance']={}
gang_info['Lance']['gang']='Jets'
gang_info['Lance']['age']='20s'
gang_info['Lance']['education']='J.H.'
gang_info['Lance']['marital status']='Married'
gang_info['Lance']['occupation']='Burglar'

gang_info['George']={}
gang_info['George']['gang']='Jets'
gang_info['George']['age']='20s'
gang_info['George']['education']='J.H.'
gang_info['George']['marital status']='Divorced'
gang_info['George']['occupation']='Burglar'

gang_info['Pete']={}
gang_info['Pete']['gang']='Jets'
gang_info['Pete']['age']='20s'
gang_info['Pete']['education']='H.S.'
gang_info['Pete']['marital status']='Single'
gang_info['Pete']['occupation']='Bookie'

gang_info['Fred']={}
gang_info['Fred']['gang']='Jets'
gang_info['Fred']['age']='20s'
gang_info['Fred']['education']='H.S.'
gang_info['Fred']['marital status']='Single'
gang_info['Fred']['occupation']='Pusher'

gang_info['Gene']={}
gang_info['Gene']['gang']='Jets'
gang_info['Gene']['age']='20s'
gang_info['Gene']['education']='COL.'
gang_info['Gene']['marital status']='Single'
gang_info['Gene']['occupation']='Pusher'

gang_info['Ralph']={}
gang_info['Ralph']['gang']='Jets'
gang_info['Ralph']['age']='30s'
gang_info['Ralph']['education']='J.H.'
gang_info['Ralph']['marital status']='Single'
gang_info['Ralph']['occupation']='Pusher'


gang_info['Phil']={}
gang_info['Phil']['gang']='Sharks'
gang_info['Phil']['age']='30s'
gang_info['Phil']['education']='COL.'
gang_info['Phil']['marital status']='Married'
gang_info['Phil']['occupation']='Pusher'

gang_info['Ike']={}
gang_info['Ike']['gang']='Sharks'
```

```
gang_info['Ike']['age']='30s'
gang_info['Ike']['education']='J.H.'
gang_info['Ike']['marital status']='Single'
gang_info['Ike']['occupation']='Bookie'

gang_info['Nick']={}
gang_info['Nick']['gang']='Sharks'
gang_info['Nick']['age']='30s'
gang_info['Nick']['education']='H.S.'
gang_info['Nick']['marital status']='Single'
gang_info['Nick']['occupation']='Pusher'

gang_info['Don']={}
gang_info['Don']['gang']='Sharks'
gang_info['Don']['age']='30s'
gang_info['Don']['education']='COL.'
gang_info['Don']['marital status']='Married'
gang_info['Don']['occupation']='Burglar'

gang_info['Ned']={}
gang_info['Ned']['gang']='Sharks'
gang_info['Ned']['age']='30s'
gang_info['Ned']['education']='COL.'
gang_info['Ned']['marital status']='Married'
gang_info['Ned']['occupation']='Bookie'

gang_info['Karl']={}
gang_info['Karl']['gang']='Sharks'
gang_info['Karl']['age']='40s'
gang_info['Karl']['education']='H.S.'
gang_info['Karl']['marital status']='Married'
gang_info['Karl']['occupation']='Bookie'

gang_info['Ken']={}
gang_info['Ken']['gang']='Sharks'
gang_info['Ken']['age']='20s'
gang_info['Ken']['education']='H.S.'
gang_info['Ken']['marital status']='Single'
gang_info['Ken']['occupation']='Burglar'

gang_info['Earl']={}
gang_info['Earl']['gang']='Sharks'
gang_info['Earl']['age']='40s'
gang_info['Earl']['education']='H.S.'
gang_info['Earl']['marital status']='Married'
gang_info['Earl']['occupation']='Burglar'

gang_info['Rick']={}
gang_info['Rick']['gang']='Sharks'
gang_info['Rick']['age']='30s'
gang_info['Rick']['education']='H.S.'
gang_info['Rick']['marital status']='Divorced'
gang_info['Rick']['occupation']='Burglar'

gang_info['Ol']={}
```

```
    gang_info['Ol']['gang']='Sharks'
    gang_info['Ol']['age']='30s'
    gang_info['Ol']['education']='COL.'
    gang_info['Ol']['marital status']='Married'
    gang_info['Ol']['occupation']='Pusher'

    gang_info['Neal']={}
    gang_info['Neal']['gang']='Sharks'
    gang_info['Neal']['age']='30s'
    gang_info['Neal']['education']='H.S.'
    gang_info['Neal']['marital status']='Single'
    gang_info['Neal']['occupation']='Bookie'

    gang_info['Dave']={}
    gang_info['Dave']['gang']='Sharks'
    gang_info['Dave']['age']='30s'
    gang_info['Dave']['education']='H.S.'
    gang_info['Dave']['marital status']='Divorced'
    gang_info['Dave']['occupation']='Pusher'

    return gang_info

def get_units(decay_rate=0.1,internal_strength=0.1,external_strength=0.4):

    gang_info=get_gang_info()

    names=gang_info.keys()
    names.sort()
    members=[s.upper() for s in names]

    education=[]
    for p in gang_info:
        if gang_info[p]['education'] not in education:
            education.append(gang_info[p]['education'])
    education.sort()

    age=[]
    for p in gang_info:
        if gang_info[p]['age'] not in age:
            age.append(gang_info[p]['age'])
    age.sort()

    occupation=[]
    for p in gang_info:
        if gang_info[p]['occupation'] not in occupation:
            occupation.append(gang_info[p]['occupation'])
    occupation.sort()

    marital_status=[]
    for p in gang_info:
        if gang_info[p]['marital status'] not in marital_status:
            marital_status.append(gang_info[p]['marital status'])
    marital_status.sort()

    gang=[]
```

```
    for p in gang_info:
        if gang_info[p]['gang'] not in gang:
            gang.append(gang_info[p]['gang'])
    gang.sort()

    #=============== make the network

    units={}

    for cluster in
[names,members,education,occupation,marital_status,gang,age]:
        for label in cluster:
            units[label]=Unit(label)

        if not cluster is names:
            connect_cluster(units,cluster)

    for name in gang_info:
        member=name.upper()

        connect(units[name],units[member])  # connect to the member

        for info in gang_info[name]:
            label=gang_info[name][info]
            connect(units[label],units[member])  # connect to the member

    for name in units:
        units[name].decay_rate=decay_rate
        units[name].internal_strength=internal_strength
        units[name].external_strength=external_strength

    return units

if __name__=="__main__":

    # example dynamics

    units=get_units()  # get all of the units

    reset(units)

    watch=['Art','ART','Clyde','MIKE','20s','40s']
    units['Art'].external_input=1.0

    t=0
    while t<80:
        update(units)
        t=t+1

    clf()
    for w in watch:
        units[w].plot_activation()

    legend(watch)
    title('Activating Name "Art"')
```

Appendix C- Train Proabilties.py Code

```
from utils import *
from Gang_Network import *
from Memory import *

def plot_weights(units):
    global _pc


    keys=units.keys()
    keys.sort()
    N=len(keys)
    im=zeros((N,N))
    for i,key1 in enumerate(keys):
        u1=units[key1]

        for w,u2 in zip(u1.w,u1.i):
            j=keys.index(u2.label)

            im[i][j]=w

    if _pc is None:
        ax=gca()
        labels=[]
        for key in keys:
            labels.append(''.join([x[0] for x in key.split()]))
        ax.set_xticks(arange(len(keys)))
        ax.set_xticklabels(labels,size=12)
        ax.set_yticks(arange(len(keys)))
        ax.set_yticklabels(labels,size=12)
        #pcolor(im)
        ax.hold(True)
        _pc=imshow(im,interpolation='nearest')
    else:
        _pc.set_data(im)
    #title('%.1f-%.1f' % (im.min(),im.max()))

    show()
    draw()

    return im


_pc=None


# BRI
N_stimulus=4
N_actions=3
N_hidden=6

world=[]
for i in range(N_stimulus):
    world.append('Stimulus %d' % (i+1))
```

```
print world


ta=[]
for i in range(N_actions):
    ta.append('Teacher Action %d' % (i+1))

print ta

ma=[]
for i in range(N_actions):
    ma.append('My Action %d' % (i+1))

print ma

h=[]
for i in range(N_hidden):
    h.append('Hidden %d' % (i+1))


print h


#raw_input('Hit return to continue')

units={}
for cluster in [world,ta,ma,h]:
    for label in cluster:
        units[label]=Unit(label)

    connect_cluster(units,cluster)


for cluster in [world,ta,ma]:
    for label in cluster:
        for hidden in h:
            connect(units[label],units[hidden],rand()*.1)

# BRI
action_map=[[1,4,1],[2,2],[3,3]]


reset(units)
for name in units:
    units[name].learning_rate=0.5

try:
    for k in range(30000):
        print k
        for action_pair in action_map:
            action=action_pair[-1]  # last one
            stimulus=action_pair[:-1] # all but the last one

            reset_external_input(units)
```

```
            names=[]
            teach_name='Teacher Action %d' % action
            my_name='My Action %d' % action

            # BRI (probabilities of external input being 1.0)
            for s in stimulus:  # allows for more than one stimulus unit
                stim_name='Stimulus %d' % s
                units[stim_name].external_input=1.0 if rand()<.15 else 0
                names.append(stim_name)

            units[my_name].external_input=1.0 if rand()<.1 else 0
            units[teach_name].external_input=1.0 if rand()<.1 else 0


            names.append(teach_name)
            names.append(my_name)

            if k==0:  # print out the training data
                print "Units stimulated: ",
                print ", ".join(names)


            # this part does the actual learning
            # BRI: may need to be longer
            for i in range(1):
                for j in range(100):
                    update(units)

                update_weights(units)
                im=plot_weights(units)
except KeyboardInterrupt:
    print "Stopping"
finally:
    print "Saving."
    reset(units,keep_weights=True)
    for name in units:
        units[name].learning_rate=0.0


    Remember(units)    # save the units to a file
```

Appendix D- Test.py Code

```
from utils import *
from Gang_Network import *
from Memory import *

units=Remember()  # recall the units from the file

activate='My Action 1'
#activate='Stimulus 1'
#activate='Teacher Action 3'
units[activate].external_input=1.0

t=0
while t<80:
    update(units)
    t=t+1

watch=[w for w in units.keys() if 'My Action' in w]
figure(1)
clf()
for w in watch:
    units[w].plot_activation()

legend(watch)
title(activate)

watch=[w for w in units.keys() if 'Teacher Action' in w]
figure(2)
clf()
for w in watch:
    units[w].plot_activation()

legend(watch)
title(activate)

watch=[w for w in units.keys() if 'Stimulus' in w]
figure(3)
clf()
for w in watch:
    units[w].plot_activation()

legend(watch)
title(activate)

watch=[w for w in units.keys() if 'Hidden' in w]
figure(4)
clf()
for w in watch:
    units[w].plot_activation()

legend(watch)
title(activate)

show()
```

REFERENCES

Baars, B. J., & Laureys, S. (2005). One, not two, neural correlates of consciousness. *TRENDS in Cognitive Sciences* , 269.

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks.* Malden: Blackwell Publishing.

Block, N. (1996). How can we find the neural correlate of consciousness. *TRENDS in neuroscience* , 456-459.

Block, N. (2005). Two neural correlates of consciousness. *TRENDS in Cognitive Sciences* , 46-52.

Bodovitz, S. (2008). The neural correlate of consciousness. *Journal of Theoretical Biology* , 1-5.

Borenstein, E. R. (2005). The evolution of imitation and mirror neurons in adaptive agents. *Cognitive Systems Research* , 229-242.

Buccino, G., Binkofski, F., Fink, G., Fogassi, L., Gallese, V., Seitz, R., et al. (2001). Short Communication: Action Observation activates premotor and parietal areas in somatotopic manner: an fMRI study. *European Journal of Neuroscience* , 400-404.

Carey, B. (2008, December 22). Blind, Yet Seeing: The Brain's Subconscious Visual Field. *New York Times* .

Cleeremans, A. (2008). Consciousness: the radical plasticity thesis. *Progress in Brain Research* , 19-33.

Crick, F., & Koch, C. (2003). A framework for Consciousness. *nature neuroscience* , 119-126.

David, N., Newen, A., & Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition* , 523-534.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor Facilitation During Action Observation: A magnetic Stimulation Study. *Journal of Neuroscience* , 2608-2611.

Fell, J., Elger, C. E., & Kurthen, M. (2004). Do neural correlates of consciousness cause conscious states? *Medical hypotheses* , 367-369.

Foreknowledge and Free Will (2008). *Stanford Encyclopedia of Philosophy* .

Gallace, A., & Spence, C. (2008). The cognitive and neural correlates of "tactile consciousness": A multisensory perspective. *Consciousness and Cognition* , 370-407.

George, D. (2008). *How the Brain Might Work: A Heirarchical and Temporal Model for Learning and Recognition.*

Greenfield, S., & Koch, C. (2007, October). How Does Consciousness Happen. *Scientific American* , pp. 76-83.

Hauser, M., Cushman, F., Young, L., Jin, R. K.-C., & Mikhail, J. (2007). A Dissociation Between Moral Judgements and Justifications. *Mind & Language* , 1-21.

Hawkins, J. (2005). *On Intelligence.* New York: Holt Paperbacks.

Jacobs, J., & Broyde, I. (2002). *Free Will*. Retrieved July 12, 2008, from Jewish Encyclopedia: http://www.jewishencyclopedia.com

Jaynes, E.T. (2003). *Probability Theory: The Logic of Science.* New York: Cambridge University Press.

Jaynes, J. (1990). *The Origin of Consciousness in the Breakdown of The Bicameral Mind.* Boston: Houghton Mifflin Company.

John, E. (2001). A Feild Theory of Consciousness. *Consciousness and Cognition* , 184-213.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *TRENDS in Congnative Sciences* , 501-507.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing Sounds, understanding actions: Action representation in mirror neurons. *Science* , 846-848.

Lau, H. (2008). A higher order Baysian decision theory of Consciousness. *Progress in Brain Research* , 35-48.

Machamer, P., & Sytsma, J. (2007). Neuroscience and Theoretical Psychology: What's to Worry About? *Theory & Psychology* , 199-216.

Maher, M. (1909). *Catholic Encyclopedia: Free Will*. Retrieved July 12, 2008, from New Advent: http://www.newadvent.org/cathen/06259a.htm

McGeoch, P. D., Brang, D., & Ramachandran, V. (2007). Apraxia, metaphor, and mirror neurons. *Medical Hypothesis* , 1165-1168.

Miller, K. (1999). Finding Darwin's God. *Brown Alumni Magazine* .

Nava, A. S. (2007). Empathy and Group Analysis: An Integrative Approach. *Group Analysis* , 13-28.

Nine Beliefs of Hinduism. (2008). Retrieved July 12, 2008, from Himalayan Academy: http://www.himalyanacademy.com/basics

Oztop, E. (2007). *Models of mirror system.* Retrieved November 13, 2008, from Scholarpedia: http://scholarpedia.com/article/Models_of_mirror_system

Vedantam, S. (2007, May 28). If It Feels Good to Be GOod, It Might Be Only Natural. *Washington Post* .

Wade, N. (2007, September 18). Is 'Do Unto Others' Written Into Our Genes? *The New York Times* , p. F.1.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *The Royal Society* , 593-602.