Monkey see monkey do: A cognitive model to loss of free will

Pars Mutaf (pars.mutaf@ege.edu.tr)
Ege University, Izmir
Turkey

Abstract

Free will (if any) may be lost because of fear and our dependence on others. In this paper, we develop cognitive models to support this argument. We do not describe what free will is but what free will is not. Our results indicate that moral responsibility is reversed and society is at fault. We argue that judgment (by society) is similar to pouring gasoline on fire, and forgiveness is the only way to regain free will if any.

1 Introduction

Free will is our apparent ability to make choices freely. The question of free will has important social implications. For example, it may hold implications for whether individuals can be held morally accountable for their actions [1].

The question of free will remains one of the most important questions in science and philosophy. Quantum mechanics, on the other hand, predicts events only in terms of probabilities, rising doubt on whether the universe is deterministic. This suggests that although our behavior is impossible to predict perfectly in practice, free will is just an illusion [1]. For decades scientists attempted to develop mathematical models of the mind, in order to understand how it works, develop a machine that is more intelligent than the human brain, and to seek answers to some important questions e.g. the question of free will [2] [3]. Neuroscientists discovered that a person's brain commits to certain decisions before the person becomes aware of having made these decisions. These findings may, however, not be sufficient to contradict free will [4]. Recent research result showed that a "butterfly effect" in the brain makes the brain intrinsically unreliable, providing the individual an excuse when their mind plays "tricks", supporting that they are not responsible for their behavior [5].

In this paper, we offer a different contribution to the question of free will. We realized that mystics and psychologists have important arguments concerning free will. We also bring our own arguments and present cognitive models that help understand how free will can be lost. We do not attempt to explain what free will is, however.

2 What free will isn't?

Mystics and psychologists noticed what free will is not, although they do not explain what free will is (see [6] and [7] for samples of the mystic and scientific argument respectively). They believe that we loose free will because mind is an illusion, an outcome of years of past conditioning. Mystics famously quote this as: *You are not your mind*. When we make a decision in our life, although we believe that it is our decision, our mind that is making the decision is actually an outcome of our previous life experiences over which we have no control. Life made the decision not us.

Mind (or, Ego [8]) compulsively thinks in order to be assured of its future existence and self-esteem, rather than simply knowing its own self and the present. It is acting out of fear and past conditioning, i.e. not free will. Thinking of the past brings back past conditioning and leads to illusory behavior. Thinking of the future is also mostly useless and triggers fear again leading to illusion. *Now* is the only reality that we have and staying in the now, i.e. avoiding thinking about past and future, sets us free from illusion according to [6].

3 A cognitive model: Monkey see monkey do

Inspired from the above arguments, we have developped the cognitive model presented in Figure 1. We model the thought process as a sequence of thoughts where the next thought t_{i+1} depends on the current thought t_i, input from other persons via five senses (with a weight $0 \leq W \leq 1$), past experiences and emotions. Another important input may be body chemistry. For example, it was shown that by manipulating serotonin levels, a causal link between serotonin and aggressive responses can be observed [9]. During years of life experience the person has billions of thoughts which are influenced by input from external world. These thoughts and inputs form what we call conceptions and misconceptions. These can be our own ideas, ideas we learned and accepted from others without questioning, behaviors that we formed our selves or copied from others, etc. When we have no time to think, when we fear, do not know how to behave, we show learned behavior, i.e. apply one of the registered conceptions. We may also show learned behavior because we no have a better replacement, i.e. do not know how to behave. We may adopt and apply these conceptions with some probability m.

We may also create our own conceptions and give them weights. Let a a given self-created conception that has a weight W_a, and b an adopted conception with weight W_b that contradicts a. Which one we use depends on our self-esteem. We can gain our independence by carefully assigning weights to inputs from outside world, i.e. questioning everything. We can for example assign $W = 0$ by default and assigning a higher value when it is necessary. For example, when we learn rock climbing it is probably a good idea to assign our teacher a larger weight then our own conception of the present problem. This too should be done carefully, i.e. after observing their experience and checking their credentials. $W = 0$

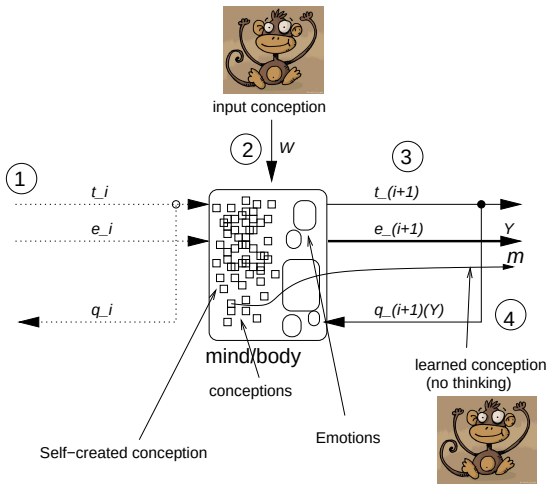


Figure 1: Monkey see monkey do (a cognitive model).

should be assigned based on topic not person. A person that was assigned a large W may provide opinions on topics they have no idea.

In Figure 1, q_{i+1} is the probability that t_{i+1} is fed back to the memory, it is a function of current emotion's intensity. If the thought is considered important, i.e. the associated emotion e_{i+1} has a high intensity Y, $q(Y)$ would be large (for example, linear increase/linear decrease).

4 Self-esteem and free will

In this section, we define two sorts of self-esteem for the purposes of this paper.

4.1 Distorted self-esteem

Self-created and adopted conceptions can be placed in a two dimensional matrix $C(t)$ as follows[1]:

$$C(t) = \begin{bmatrix} c_{0,0}(t) & c_{0,1}(t) & c_{0,2}(t) & c_{0,3}(t) & c_{0,4}(t) & \cdots \\ c_{1,0}(t) & \mathbf{c^s_{1,1}(t)} & c_{1,2}(t) & c_{1,3}(t) & c_{1,4}(t) & \cdots \\ c_{2,0}(t) & c_{2,1}(t) & c_{2,2}(t) & \mathbf{c^s_{2,3}(t)} & c_{2,4}(t) & \cdots \\ c_{3,0}(t) & c_{3,1}(t) & c_{3,2}(t) & c_{3,3}(t) & \mathbf{c^s_{3,4}(t)} & \cdots \\ c_{4,0}(t) & \mathbf{c^s_{4,1}(t)} & c_{4,2}(t) & c_{4,3}(t) & c_{4,4}(t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where $c_{i,j}$ is a conception and its value is its weight W which can change as a function of time. Among these we define the bolded ones $c^s()$ as the conceptions

[1]One or three dimensional matrices could also be used. We choose a two dimensional one for notation simplicity.

about self (adopted or self-created). Self-esteem then can be defined as this subset $C^s(t)$ of $C(t)$ where all elements except $c^s()$ are undefined (null).

For example, the individual may have fascist tendencies which are approved by their environment that were assigned a large W, resulting in illusory high self-esteem $c^s_{1,1} > 0$. However, a common sense conception $c^s_{3,3} < 0$ that fascism is bad also exists resulting in illusory low self-esteem especially if assigned a large weight. One approach to calculate the overall self-esteem here would be taking the sum of the two, however we believe that valuable information about self would be lost in this case. The proper approach is to see the self-esteem as both high and low. This is what we call *distorted self-esteem*.

The individual's life may be attracted to the fascist environment or the entity (individual or group) applying common sense. We believe that, as we argue later, this will depend on from which entity the individual receives love and respect.

4.2 Overall self-esteem

The overall self-esteem is also important. It can be calculated as the proportion of self-defined conception use rate to total conception use rate. In this case all elements in $C(t)$ are taken into account. Conception use rate is however not easy to compute. One would for example need to sample the number of conceptions (own or adopted) used per second. The following function is easier to calculate and provides more information (inspired from control theory [10]):

$$a = \begin{cases} 1 & \text{success} \\ -1 & \text{failure} \end{cases}$$

Self-esteem can be calculated as follows:

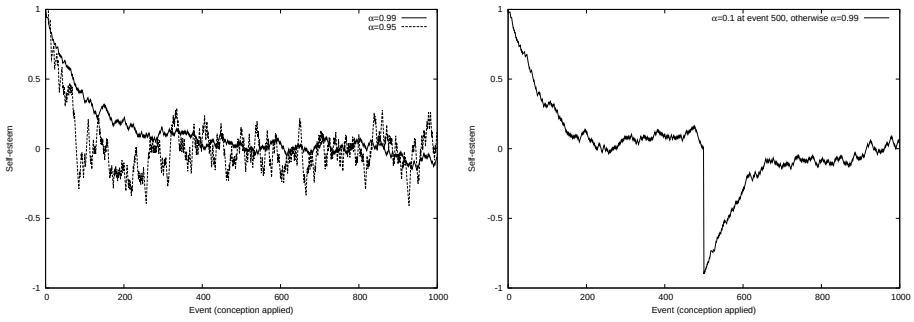
$$e_s(t) = e_s \leftarrow \alpha e_s + (1 - \alpha)a \tag{1}$$

where $-1 \leq e_s(t) \leq 1$ and $0 \leq \alpha \leq 1$. The larger α the less importance is given to the current event (success or failure), current estimation of self-esteem (cumulative result of past events) is given more importance and changes slowly. The choice α depends on the individual and the current event. If the individual gives more importance to their current self-esteem α would be large, if the current event is given more importance it would be small. The definition of success and failure needs to be done here. Considering the results of the adopted conceptions, there are the following possibilities:

1. Own conception is applied, result is good (success)
2. Own conception is applied, result is bad (?)
3. Adopted conception is applied, result is good (?)
4. Adopted conception is applied, result is bad (?)

How α is set in each case is hard to estimate, except (1) where it can be chosen large and increase their self-esteem. (2) may reduce self-esteem by causing the individual to choose a low α in response to the event, depending on the emotion associated with the result, although the individual used own conception. The effect of (3) is difficult to understand. There is success but credit goes to another. The individual may be proud to adopt a correct conception, but if they adopt too many conceptions, although correct, this is a sign of low self-esteem. (4) may have strange impacts on self-esteem. Instead of using a self-conception or applying rational thinking, the individual trusted someone else and failed. The individual may not care because it was not own conception or regret and choose a very low α. We believe that this model reflects well the self-judging nature of human being, which is done by chosen a low α. The individual sacrifices their self-esteem.

We assume here that α has a default value and when one the above events occur the individual gives it a new value, $e_s(t)$ is updated and α is assigned again its default value, that is the individual stops thinking about the event but self-esteem is lost. A more complicated model would change the default value of α each time a success or failure occurs. This would for example model an unstable behavior where bad and good events are over emphasized (by reducing alpha), as if the individual is overly affected by bad events and tries to regain self-esteem by exaggerating good events. For success and failure, assigning different values to α would probably be more accurate, however, because this bipolar behavior may not always be the case. This also leads us to believe that high self-esteem has a stability aspect. Such a complex model is not necessary for the purposes of this paper, however.



(a) (b)

Figure 2: Self-esteem simulation results.

Figure 5(a) illustrates $e_s(t)$ for different values of α for an individual with initially full then mediocre self-esteem (probability of using own or adopted conception are equal). Figure 5(b) shows a scenario where α is normally high but in response to a failure (# 500) individual choses a low α which lowers their self-esteem. The individual however forgives themself and increase α again,

healing their self-esteem.

We believe that the lower $e_s(t)$, the less free will we have. Because we apply others' conceptions which were also created by or adopted from others, forming a reverse tree rooted on the individual. Life makes the decision, not the individual. According to this model, the answer to the free will question (if we have free will) is not binary (yes or no), but a continuous one that depends on time.

5 Love and free will

In this section, we argue that unconditional love is the only way to achieve free will, if any. Again we do not explain what free will is, but how we can loose it. Free will may be lost because of our dependence on another's love. This is what we call *conditional love* and we describe it as follows:

- Loves for reward
- Rewards for love

Reward may be love, approval, protection, forgiveness, or even respect. The main cause may be fear, especially fear of low self-esteem. We love those who protect our self-esteem or those who nourish our self-esteem. Our thoughts and hence our life is attracted towards them, i.e. we have no free will. Those who have low self-esteem may obviously apply conditional love to survive or support their self-esteem. Those who have a false sense of self-esteem may also apply the same strategy. They may in fact have low self-esteem and need others' love and respect to gain self-esteem. They can obtain it through conditional love (i.e., I love and protect you if you respect me). Narcissism for example may be due to or accompanied by low self-esteem [11], also described as craving attention and admiration [12]. It is interesting to note that although they seem superior, they also love for reward which is respect or admiration. If they do not obtain the expected reward they may apply aggressive strategies however, which narcissisic individuals apparently do [11].

Unconditional love on the other hand does not need others to gain self-esteem. This requires however letting go all fear. A person with a healthy self-esteem accepts and loves themself unconditionally, acknowledging both virtues and faults in the self, and yet, in spite of everything, being able to continue to live loving themself. They show minimal aggressive behavior because they have high and stable self-esteem and their self-love remains the same no matter what happens [13]. Unconditional love is not passively being victim of others, it is however total forgiveness. Unforgiveness is being dependent on those that we do not forgive, because we are basically expecting them undo the harm which they cannot or may not be willing to. Conditional love, loves in order to self-protect from harm (reward being here not being harmed). Unconditional love, on the other hand, does not fear being harmed however is always careful for themself and others.

Unconditional love is different from maternal or romantic love. Scientists also predict that since unconditional love experientially differs to a large extent from romantic love and maternal love, this form of love would be mediated by brain regions not involved in romantic love and maternal love [14]. Unconditional love is also different from compassion. Compassion refers to an awareness of the suffering of another coupled with the desire to alleviate that suffering. In contrast to compassion, unconditional love is not specifically associated with suffering [14].

Human being is capable of both unconditional love and conditional love as illustrated in Figure 3. Unconditional love is our best state, we have no loss of free will because our mind and actions do not depend on others. With fear we begin to descend since we become dependent on others and loose free will. When the conditions are always satisfied, i.e. we receive a reward when we love, or we are loved when we reward, we may be lost in delusion and feel happy and achieve a very high conditional love state. This is in fact an illusion. Although we feel happy, we became highly dependent on others and lost our free will. We model this transition as a U (or S) curve since we estimate that the illusion, i.e. transition to conditional love state at the break point $x = 0.5$ would be a smooth one. The individual may become dependent without realizing it.

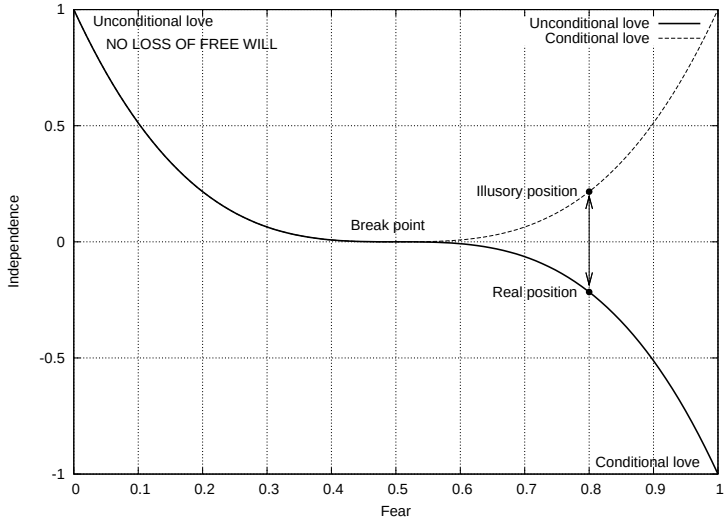


Figure 3: Unconditional versus conditional love.

It is interesting to note that independence takes positive and negative values, or zero. When it is 1 we are fully independent and our life is led by our own free will. When it is zero, we have no free will but we are not dependent neither. We believe that plants can be the closest example that we can have at this state. Not all plants are at this state because some of them needs humans to survive. Humans on the other hand may fall further becoming dependent on

other humans, i.e. living according to their will (hence, negative independence). We believe however that where we are on this curve depends on circumstances. It may very rapidly change and consequently it is impossible to label a person as dependent or independent. For example, a murderer may save a random child putting in danger their own life, depending on the circumstances.

We assume that independence has maximum and minimum values, 1 and -1. We also assume that when our fear reaches a maximum value 1, we become most dependent and are trapped by conditional love. This curve can be modeled as:

$$y = -(2x - 1)^3$$

In this case, we can define W as a function of independence as follows:

$$W = -0.5(y - 1)$$

The more dependent we are with another individual, the larger the weight we assign them ($0 \leq W \leq 1$).

6 Society at fault?

In this section we try to find out what moral responsibility is in the light of above models. We assume the existence of two entities:

1. Society
2. Dark Side (DS)

Dark side may be any entity e.g. a person, a group, or concept, that is considered faulty by the society. We assert the existence of the *vicious loop* illustrated in Figure 4. The individual commits a fault, is judged (meaning rejected) by the society, they are attracted to the DS which gives them love (although conditional), causing distorted self-esteem (both low and high), in need of self-esteem and love (even protection from society), they adopt another dark conception from DS and commit another fault. The probability of one round is $a \times b \times c \times d$. The individual may begin with little faults, which become more and more important with time until themself become a DS. It is important to note that the probabilities a, b and d depend on how much love is received from DS and how intensely they are rejected by the society. The probability c depends on how unforgiving the society is.

One can argue that the DS or the individual are at fault, however society also is in a very suspect position, we because they are constantly pouring gasoline on fire by rejecting the individual. Assuming that DS will change would put the society in a weak position. Consequently, society must change. Thus, moral responsibility is forgiving and teaching self-forgiveness as soon as possible.

This does not mean that we have no moral responsibility. A person who is constantly at fault should be stopped to protect others. However, there is currently no proof that for example putting individuals who have been at fault

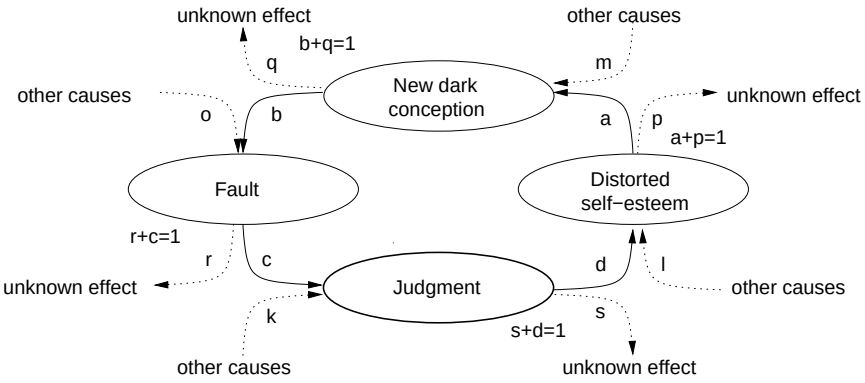
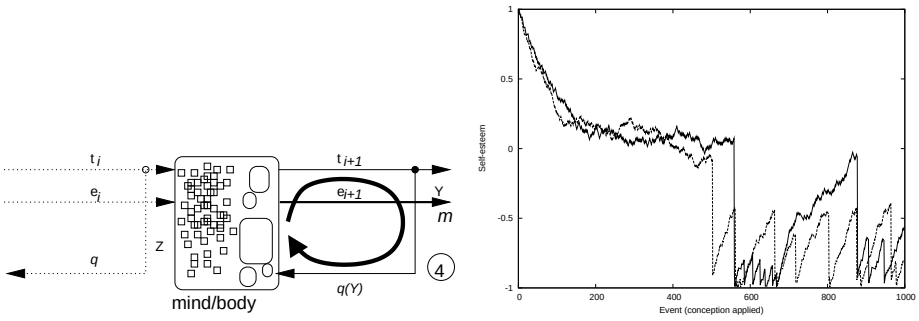


Figure 4: The probability that an individual is trapped in the vicious loop.

in jail is beneficial for them or society. According to our model, it is impossible to judge them and healing them via teaching self-forgiveness may be a better approach. Humanity, however, is probably not yet at this stage. Achieving such a level of forgiveness, i.e. forgiving atrocities, requires great courage and strength. We refuse to forgive faulty behavior because we fear. We accept others judgment for the same reason. Only forgiveness can break this loop, which requires letting go all fear. Unconditional love, loves self and others equally. The individual has high esteem and does not adopt random conceptions from others, reducing the fault rate. Unconditional love does not judge, which breaks the above loop. Faults need to be avoided, however according to the above model judgment and punishment is not the ideal solution. A better approach may be: *This is wrong, but not your fault. This is not your nature* (i.e. without distorting their self-esteem).



(a) (b)

Figure 5: Self-judgment simulation results.

In support of this argument, Figure 5(a) shows how our model reflects the effect of judgment, which is a strong emotion. The individual commits a fault

and is judged by another that were assigned a large W. Being judged by a loved one creates a strong emotion, in which case Y, and hence q is large and we are certain that the problematic thought comes again and again. The individual may arrive at the conclusion that the problematic thought or behavior is in their nature because it is frequently experienced, distorting their self-esteem. For example, the individual is observed playing with a gun, and judged. Having low-esteem the individual constantly thinks about the gun, with some probability wondering if they are attracted to weapons (creating a high emotion) because this would be the only real reason for judgment. In fact judgment was not made for a particular reason (real reason may be fragility). An individual with high self-esteem probably would not take the event seriously. Figure 5(b) shows two simulation results using Equation 1. At event #500 the individual is judged, then with a probability 1/25 an event occurs (in their mind) which is self-judgment (adopted conception) about the event which again and again lowers the self-esteem with a probability 1/25 until they forgive themself i.e. stopping putting a high emotion Y to the event. We observe here that α is reduced because of high Y. There may also be other factors however.

7 Solution space

According to above results, we may loose free will because of fear, which lowers our self-esteem, leading to the adopting of random conception instead of applying our own thinking. In the light of these results, in this section we review possible approaches to retrieve free will.

7.1 Self-discovery

Mystics believe that God is all that is including each human, animal and plant [15]. God is modeled as an ocean where each individual is both a drop in that ocean and the whole ocean (anonymous). This is referred to as "oneness" which mystics apparently experience through meditation (see [16] for a scientific view of meditation and oneness). Thus we are God, however lost our true nature for some unknown reason. God is *unconditional love*, which apparently mystics feel in meditation [17], they also call it "love with no object", or "love for no reason". Mysticism refers to "the practice of knowing the reality of God through direct experience.", or self-discovery or (discovery of God within). Unconditional love is pure silence, absence of all mind distortions (i.e. misunderstandings) about love [18].

7.2 Power of Now

Mystics realized that fear and self-esteem issues reside in the past and future hence the developed a technique that consists of staying in the now. The technique consists of watching the thoughts without judging them. Mystics suggest becoming intensely conscious of the present moment, without thinking about it

or labeling its contents. This creates a gap in the mind's thought stream and also awareness of one's own presence. Any routine activity, such as walking up stairs, or washing one's hands, can become a vehicle for present-moment awareness. They pay close attention to every step, every movement, even their breathing. By watching the thoughts and staying in the now, one can realize that they have less thoughts. One can let go old mind habits, i.e. past conditioning which are certainly not free will. By thinking less they leave space to intuition.

Quoting [6]:

"The moment you start watching the thinker, a higher level of consciousness becomes activated. You then begin to realize that there is a vast realm of intelligence beyond thought, that thought is only a tiny aspect of that intelligence."

8 Conclusion

Free will (if any) may be lost because of fear and our dependence on others. In this paper, we develop cognitive models to support this argument. We do not describe what free will is but what free will is not. Our results indicate that moral responsibility is reversed and society is probably at fault. We argue that judgment (by society) is similar to pouring gasoline on fire, and forgiveness is the only way to regain free will if any.

We believe that mystics may be very much advanced on the problem and cited references for future work. The mystical argument is: *even if free will exists it requires hard work*. This includes traditional techniques e.g. meditation, or more modern ones like watching the mind/body and staying in the now.

References

[1] Wikipedia, "Free will `http://en.wikipedia.org/wiki/Free_will`," 2012.

[2] Leonid Perlovsky, "Free will and advances in cognitive science," *arXiv:1012.3957v1*, 2010.

[3] Leonid Perlovsky, "Toward physics of the mind: Concepts, emotions, consciousness, and symbols," *Physics of Life Reviews*, vol. 3, no. 2355, 2006.

[4] C. S. Soon, M. Brass, H. Heinze, and J. Haynes, "Unconscious determinants of free decisions in the human brain," *Nature neuroscience*, vol. 11, no. 5, pp. 543545, 2008.

[5] M. London, A. Roth, L. Beeren, M. Husser, and P. E. Latham, "Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex," *Nature*, vol. 466, no. 7302, 2010.

[6] Eckhart Tolle, *The Power of Now: A Guide to Spiritual Enlightenment*, Hodder & Stoughton, 2005.

[7] A. Laitinen, "Charles Taylor and Paul Ricoeur on self-interpretations and narrative identity," *Narrative research: Voices of teachers and philosophers*, pp. 57–76, 2002.

[8] Wikipedia, "Ego (spirituality) `http://en.wikipedia.org/wiki/Ego_(spirituality)`," 2012.

[9] M. J. Crockett, L. Clark, G. Tabibnia, M. D. Lieberman, and T. W. Robbins, "Serotonin modulates behavioral reactions to unfairness," *Science*, vol. 320, no. 5884, 2008.

[10] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, 2005.

[11] B. J. Bushman and R. F. Baumeister, "Threatened Egotism, Narcissism, Self-Esteem, and Direct and Displaced Aggression: Does Self-Love or Self-Hate Lead to Violence?," *Journal of Pea-sonalityand Social Psychology*, vol. 75, no. 1, pp. 219–229, 1998.

[12] C. C. Morf and F. Rhodewalt, "Unraveling the Paradoxes of Narcissism: A Dynamic Self-Regulatory Processing Model," *Psychological Inquiry*, vol. 12, no. 4, pp. 177196, 2001.

[13] Wikipedia, "Self-esteem `http://en.wikipedia.org/wiki/Self-esteem`," 2012.

[14] B. Beauregarda, J. Courtemanchea, V. Paquettea, and . L. St-Pierrea, "The neural basis of unconditional love," *Psychiatry Research: Neuroimaging*, vol. 172, no. 2, pp. 93–98, May 2009.

[15] Neale Donald Walsch, *Conversations With God : An Uncommon Dialogue (Book 1 and 2)*, G.P. Putnam's Sons, 1995.

[16] Glen Peter Kezwer, *Meditation Oneness and Physics*, Steiner Books, 2003.

[17] Harriet Mohr, *The God Within: Meditations*, New FOcus Press, 2004.

[18] Silence, "Pure silence `http://www.puresilence.org`," 2012.