Doozandeh, P. (2014). *Emotion behind intelligence* [unpublished manuscript]. http://p-doozandeh.com

Emotion Behind Intelligence

Pooyan Doozandeh

Author note:

This manuscript was completed in July 2014, and uploaded online in February 2021.

No funding agency or university supported the creation of this manuscript. Correspondence concerning this manuscript should be directed to Pooyan Doozandeh (pooyan.doozandeh@gmail.com).

Abstract

After about sixty years of the activity of researchers and scientists in artificial intelligence, there is not even a trace of applying human intelligence and its capabilities into computers. It seems that researchers forget about the grandiose promises of the founding fathers of AI and the highest end of this field which is to build machines as intelligent as human. There is a common perspective which considers the mind in terms of computation and representation that is not only in AI but in other fields like cognitive science. In this article I claim that because human emotion is an integral part of intelligence, and because building computers with emotions or the computer modeling of mind with emotions is impossible, the idea of having human intelligence in computers is doomed. Human intelligence and mental activities are completely influenced by human emotion and emotion is nothing but the manifestation of living organism's desire for survival and reproduction.

Keywords Emotion, AI, Cognitive Science, Affective Computing

Introduction

Emotion plays an integral role in every aspect of human intelligence. There are unanimous voices verifying the importance of emotion not only in philosophy but more recently in scientific discussions (see Damasio 1994). This essay identifies the emotion as the reason of stagnation in the realization of promises in artificial intelligence and cognitive science. By promise I mean both original humanoid intelligent systems and more recent applied oriented AI and cognitive science like natural language processing and expert systems. I first discuss inherent problems with active areas of artificial intelligence and cognitive science with regard to their original goals and I believe that although there have been tangible benefits of these fields in many areas of technology, they will never get to their highest goal; at least with current perspective. Then I will mention Affective Computing as a field that truly recognized the importance of emotion in intelligence but has focused only in recognition and expression of emotions. I bring evidence for my claim of the primacy of emotion in human intelligence from scientific and neurological literature.

Artificial Intelligence

Irrationality of Claims

The people who claim the possibility of traditional Artificial Intelligence believe that arguments against the basic premises of AI must show that building a working AI system is impossible. But here I do not take this presupposition for granted since this method has failed to convince the traditionalists of AI of its impossibility. Instead I will suppose that AI has finally succeeded in its promise to show how implausible the promise looks after its realization.

Pioneers of the Artificial Intelligence had this notion that by working in the field, we will be able to build an artificial mind that has the exact capabilities of a human mind, but in my opinion they were not aware of the implications of their belief. If we accept that, by using our intelligence we can make an artificial mind that has the same capabilities of human mind, then the machine is as intelligent as we are and we can expect the machine to have the human reasoning power. Those who assert such argument also accept that human uses the same reasoning power in areas like mathematics and physics and virtually every discipline of science and technology. So if machines can have the same intelligence and reasoning capacity of us,

then we can expect them to be mathematicians or physicians or in general they can take the place of every scientist or technician.

But in my view, there is a logical contradiction in that assertion. It is clear that every machine is the product of technology. In logical term, it is the effect of technology. Now, let's presume that we built an AI that is as intelligent as us. It means that this AI has the power to cause and expand the technology the way that humans have done. The problem is here: if technology causes AI then it is impossible for AI to cause or alter its cause since in logic, cause precedes effect and there is no means by which the effect causes its cause. In a simpler notational form:

$$((\text{Technology}) \longrightarrow (\text{AI})) \quad \not\longrightarrow \quad ((\text{AI}) \longrightarrow (\text{Technology}))$$

So, if we accept the first part, we can never expect the second part. Also note:

$$(\text{Human Intelligence}) \longrightarrow (\text{Technology}) \longrightarrow (\text{Artificial Intelligence})$$

$$\neq$$

I make another simple logical argument. Regardless of the conclusion above, again imagine that the AI scientists have finally succeeded in building an AI which is comparable with human intelligence. Because we used our intelligence to make an intelligent machine, it is not rationally impossible to expect the intelligent machine to build another exact copy of itself. So the machines can clone another machine by the use of their intelligence. The problem is if that is possible, we proclaim a metaphysical and supernatural rule that intelligent beings can make another copy of an intelligent being just with their intelligence and then we can expect the humans or even animals to make an exact copy of themselves by the use of their intelligence. In fact it is absurd to wish humans or animals to clone themselves not biologically and asexually by the use of reason and intelligence.

Another point that needs to be clear is that when AI claims to implement human intelligence into machines, whose intelligence is it? They have entirely ignored the differences between human individuals, but I see noticeable differences in their intelligence. Even when it comes to use logic and reason there are differences between individuals. The way that I use my mind to solve a mathematical problem is probably different from another person's way. The problem is that I can never know what processes the other person's mind go through to solve that problem.

The arguments that I made, although seem simple, has been ignored by pioneers and visionaries of Artificial Intelligence because they focus only on their claim of possibility of building an intelligent machine. But they have never thought about the logical consequences of their claim after they come through and their discussion has always been about the time before building an intelligent machine.

The Problem with Distinction

Another topic that I am interested to analyze is about a perennial mistake of many people in the entire history of AI. Some still think about the human mind as an entity which is composed of distinct and separate parts: reason, consciousness, memory, emotion, and so on. So far many philosophers have tried to prove that having human reason without consciousness is unimaginable; but because the definition of consciousness is subject to many diverse interpretations, I will continue with discussion on the human emotion, for I think is the dominant force behind human intelligence.

In my view considering the mind with such distinct parts is invalid and it is wrong to expect the human mind to operate as only a reasonable system, because I think that these parts are so tightly coupled together and their operation forms a holistic entity. I make an argument which is the primary subject of this essay: the mental components like reason, memory, and consciousness are closely interconnected and one of the most important inseparable factors behind all of them is the dominant force of emotion. To explain this view I will emphasize on the pervasive dominance of emotion in every facet of human existence and of course reasoning ability. I will discuss it in details in incoming sections of this essay but it is enough to mention that human uses his reason in every problem or domain that his emotion leads him to. Imagine a physicist who uses his reasoning power to prove a theory. If he does not have any positive feeling in what he is doing, he will never be able to work on his theory. No one can deny that

every activity of ours is the result of the dominion of emotion and the reasoning ability is not an;2 exception. Here I quote a remarkable statement of David Hume in his *Treatise of Human Nature:* "Reason is and ought only to be the slave of passions; the only work it can claim to do is in serving and obeying them". [1]

Today's Artificial Intelligence

I will try to show the influence of emotions in the constitution of the common notion of human mind. So if AI scientists claim that they will be able to build an Artificial Intelligence with the abilities of human mind, they have to find a way to simulate the human emotion. This section discusses the most active fields of today's AI and I will buttress my argument by showing how these fields are in stagnation in approaching the human mental capabilities.

Decision Making

The traditional approach of artificial intelligence in understanding human decision making has so far been inspired by neo-classical economic theories which promulgate a rationalist theory of knowledge. This approach is called Rational Decision Making and was welcomed by computer and AI scientists because computers are unable to implement anything but rules of logic and reason. But a simple question is: do humans make decisions rationally?

Although in computer world, the rational decision making can solve many small-sized planning problems, but when it comes to simulate decision making in human, it is absurd to expect this rational approach to decide like human since humans are far from being rational. ;8Even if we knew to what extent human is rational, many of the most important problems in economics, politics, sociology, and other humanistic sciences would be solved.

Behaviorists had tried to prove that by knowing a person's historical environments, we can predict that person's reaction, or decision, to certain stimuli or choices, but they were unsuccessful. Even by incorporating the genetic information with a person's past and present environments, we are unable to make such a prediction.

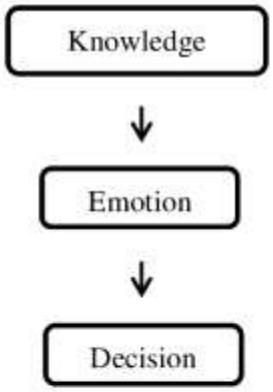
The people who follow the rational approach believe that the major reason of their failure in simulation of human decision making is that computer programs lack the human knowledge of the world and if we can represent the human knowledge to computers, we will get closer to the implementation of human decision making. Therefore they focused their efforts to another discipline which is Knowledge Representation. They argue that making a

rational software is simple; but the differences in decisions, even in human individuals, stem from the difference in their magnitude of knowledge of the world.

It is evident that differences in decisions depend, to some extent, on our knowledge of the world and more knowledge leads to better decisions (with respect to our goals), but I claim that even if AI scientists were successful in knowledge representation and computers could have the same knowledge as that of an average human, they would not be able to simulate the human decision making. These people have never considered the prominent factor behind human decision making which is emotion. In my opinion, in decision making, the emotion is the factor that filters all other effects, like knowledge, before the decisions are being made. The knowledge of the world does not influence the decisions directly; the only thing that it affects is our emotion towards different available options. It is noticeable that by acquiring more knowledge about different options, or in a broader sense everything, our feeling towards those options or things changes and when our feelings change, our option-selection changes.

Emotion is the only involving factor in decision making. In psychological terms, we tend to choose the options that we have a positive valence about. An important point is that the knowledge, however important, is not the only influential factor in our feelings towards different options. There are many other complicated and interwoven factors affecting our feelings towards things. I will discuss these factors and the essence of emotion later.

I simplify my proposed model of human decision making in this picture:



Statistical Methods

Statistical methods are used in many areas of artificial intelligence. Statistical Pattern Recognition and Machine Learning are the two conspicuous examples of this method. Let's consider statistical pattern recognition which is an active field of research in many universities.

Think about a classifier. I think everyone agrees that a classifier is a set of rules that is built upon the observation of instances in the training set. These rules try to recognize the class of future observations. So the only thing that we see is these rules trying to recognize instances and categorize them in different classes and most of the current researches in the field are centered on how to make rules with higher accuracy for future observations.

Now the simple question to ask is do humans recognize different things by a set of rules in their memory? Consider the example of recognition of an orange from an apple. When we look at an orange, we make an unconscious immediate mapping of the orange and many impressions in our memory like its taste, smell, texture, and at the same time the feelings associated whit it like whether we liked it or not is provoked in our mind. This mapping that leads to associated feelings is based on our previous experiences of oranges and more recent experiences have stronger impacts on our feelings. These steps happen in a fraction of a second and we are not fully aware of them and many other contributing factors involved. The same steps for the apple.

I believe that when we make such recognition between different things, we do not compare quantitative or qualitative values of a number of features or attributes; but for every single thing for us to recognize, there are relevant impressions and feelings that without them we are not able to recognize them and that is the reason why we cannot recognize things that we have never had any experience of.

The highest end of statistical pattern recognition is the implementation of human ability of recognizing different patterns in computers. But it seems that the researchers of the field are not eager to think about the fundamentals of their studies and involving details about the human recognition power and that is in my view the major obstacle of their progress. I will not continue on machine learning since the approaches and methods are not much different from pattern recognition and these two fields share the statistical method and they differ only in minor details like the presence or absence of supervision during training time. The primary existing factor in both areas is their feature-oriented and rule-based attitude towards

recognition and as long as this attitude persists, I do not expect much progress in approaching human recognition power.

Expert Systems

From its birth, researchers of expert systems had this notion that they could replace the human expert in every area of expertise with machines; based on the assumption that every human expertise can be completely described by a set of rules.

Expert system is a computer program that has a rule-based engine as the primary component. By translating the environmental stimuli into a binary pattern of input, there exists a specific output or response; therefore different patterns of input lead to different responses. As input pattern, this rule-based engineering system can have a simple question-answering of a human advisor to a more sophisticated motor skill of a driver and the predominant challenge of this field is how to apply the expertise as a set of rules into the engine of the system.

The question is: can we have a clear understanding of human expertise? If we defined the job of experts as solely presenting people with questions and recognizing different patterns of answers, we could easily have a simple and clear understanding of that job. But the point is that I barely associate such a job with expertise. The real expertise lies in motor-skill related activities. If we are to implement these skills into the computers, we must first think about how a person acquires such a skill.

One important reference concerning this issue is the teachings of German philosopher Martin Heidegger and its later interpretation of skillful coping by Hubert Dreyfus. To put it in brief, Heidegger's philosophy emphasizes on how in specific situations and contexts, we use tools to bring about our desired changes in our surrounding. In contrast to the detailed explanation of human skills, he had a holistic approach to our skillful usage of tools and had this idea that when we use tools, we and tools form a unique entity. If we want to delve deeper into this subject, we will have to think about the origin of technology and the nature of movements in humans and other life forms and the fact that every form of life has some kind of movements and as long as life exists, it shows itself by movement. I do not continue with this subject in that it is outside the scope of this article and instead I will discuss my view and simplify the matter with an example.

Imagine a child who is trying to learn how to ride a bicycle. In short, that is clear that at first he tries and fails and tries over and over again until he notices some progress in riding for

longer distances. He feels positively towards his progress and realizes that by further exercise he will be able to ride for longer distances. His positive feeling in learning supports his skill acquisition until he fully learns how to do it.

The important point is if it was not for the positive feeling of a person in every skill acquisition, that person could never learn that skill. It is not only about motor-related skills but in every other learning process in humans, as long as there is no good feeling, there will be no learning. That is in my opinion the major reason behind the fact that if the people are not emotionally in a good state, they will not be able to properly practice their skills and the more sophisticated the skill is, the more this problem stands out.

Because the skill of people is combined with their emotion, the detailed explanation of their skills as a set of rules becomes impossible. When an expert system designer asks a skilled person to describe his expertise in detail, the problem arises. Needless to say that people do not practice their skills by going through the tremendous details involved in performance, though at the same time, they regard all of those details in a perfect way and if we want to implement the same skill into machines or software, there is no way but to apply all of those details in form of precisely defined rules.

To put it simply, the skillful performer does not know how he does it but he does it perfectly. If he wants to talk about the details of his skill, he will certainly be confused, for the emotional core of his performance is inexplicable in words or rules. While I do not doubt that by future technological advancements, that will become possible to replace many experts like drivers, with rule-based systems, that system will never be like a human performer, for our inability to apply the emotional core of our performance.

Natural Language Processing

The possibility of having a natural conversation with computers has always been one of the most fascinating sub-goals of artificial intelligence. So far linguists have tried to understand the evolution, formation, and structure of human natural language. The most prominent hypothesis which AI scientists were eager to implement in the computer was the structural, rule-based approach to the grammar of natural languages that was proposed by Chomsky and other linguists during 50's and 60's. Here I will not discuss the grammatical aspect of human language, but I will focus on the lexical aspect and our understanding of words.

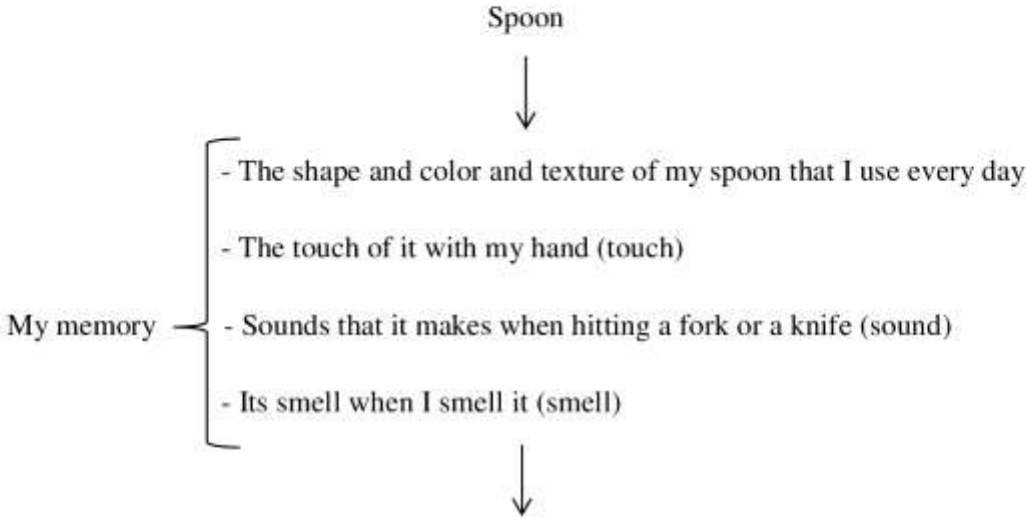
The most important factor that has again been ignored by most of the linguists and AI scientists is the impact of emotion in language acquisition of humans. Even a modicum of sensitivity will suffice for anyone to recognize that for every word in the vocabulary of a language, there is a concomitant emotional shade. In learning a language, either for a child or an adult, human mind is not a computer memory with prior structural rules. As long as emotions are not attached with a word, we will never be able to understand or learn the meaning of that word. I explain my claim in more details.

The way by which we understand words is in my opinion closely related to our recognition power. Human understanding and recognition of different words stem from the way we call the objects that we perceive or events that happen to us. This is especially more evident in childhood, since there are fewer number of abstract nouns in human mind. That is the reason why every time we encounter the noun of an object, it forms a picture in our minds and therefore at the same time, associated impressions and feelings are generated. I mentioned my arguments over human recognition in previous sections. Concisely, nouns, especially objective nouns, and visual perceptions are so closely associated I cannot make a sharp boundary between our understanding of visual objects and nouns. The same relationship exists for verbs and adjectives.

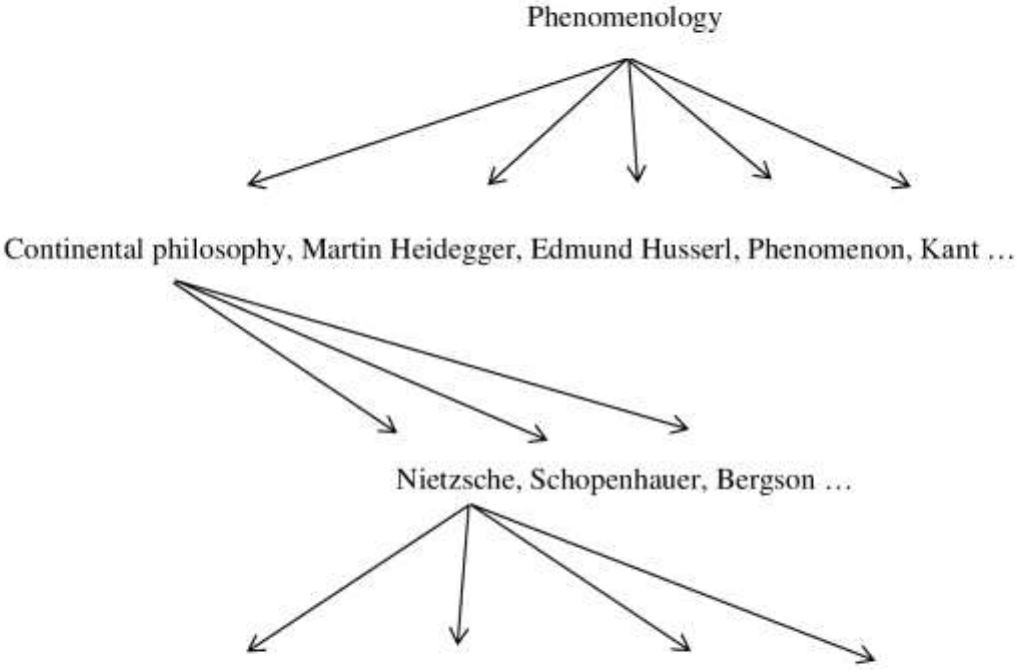
In childhood we learn simple and objective words and as we grow older, we learn more abstract words that in most cases, their meaning can be defined and analyzed by simpler words and if we follow the abstraction levels, we reach objective or simple event-related words.

In other terms, when we encounter a word, we feel some vague emotions attached to it. It seems that when we hear or read that word, these emotions affect us immediately, but when we are asked to think about the reason of those emotions, we make a short journey into our memory and analyze that word consciously. Such analysis connects us to a network of related memories that can be described by words that are closer to sensory impressions. If we continue the analysis, we would undoubtedly come to words that are used to describe specific sensory impressions that arouse specific feelings in us.

If a word is an objective one, there would be a simpler network and we can easily explain them in terms of specific feelings, but if the word is an abstract one, the network that we make in our memory to analyze it would be a large and complicated one and our journey to reach to specific sensory impressions and feelings would be longer. Here I name two words and make a simple exemplary analysis of my own memory.



Because the spoon is a simple objective word, let us think of another example of an abstract word:



Living in nature and responding to immediate desires and beyond social norms …

- Forest (picture)

- Sound of singing birds and running water (sound)

- Smell of wood (smell)

- A mild cool breeze (touch)

Associated subjective feelings of those impressions

The reason I showed these figures is because I wanted to show how thinking consciously about a word can connect us with memories of other related nouns. In the end if we are asked why we have either positive or negative valence about that word, we have to make such a journey in our memory.

Concisely, about every word there is an emotional shade and we can analyze that word in terms of emotions which is the result of specific perceptible impressions. An important point to notice is this: words that are used to describe the emotions cannot be analyzed themselves. That is, in my view, the reason that if you look up the meaning of every word in dictionaries, you will find definitions and descriptions; but if you look up the words that denote emotions, you will find no satisfactory explanation in terms of other words. You can try the words like anger, sadness, love, pleasure and what you will find is the description in more abstract words that are more misleading. As another example, the meaning of emotion in dictionaries is feeling and the meaning of feeling is emotion or a circle of limited words that try to define each other in a vague manner. That is the limit of how much we can analyze the semantic of words.

I can make the same scenario of analysis for verbs in terms of actions that transform the state of our environment and the associated feeling that we have for these actions.

In my argument for language, I agree with Wittgenstein's account of language analysis in Tractatus Logico-Philosophicus and the way he related every noun with a visual object. The difference is that he did not take the emotional side of our understanding into account and

therefore words like warmth, anger, love… were considered meaningless and were among words about which "one should be silent".

So, if there has ever been problem with the definition of our understanding of words, it is because we cannot have a clear explanation for emotions. If there is any possibility in the implementation of the human usage of language into machines, we will have to seek this possibility in the explanation and implementation of emotions.

If one accepts even some parts of my presented arguments, he will notice the strong connection between emotion and memory, as I indicated in my holistic approach to mind in the previous section of this article.

A Bit of History

If we are to know where the AI has gone wrong, it is important to find the historical assumptions which lie behind today's Artificial Intelligence. For this reason, I will briefly discuss the relevant issues in the course of the history of computer science and AI.

The history goes back to the invention of abacus and later in Charles Babbage's Difference Engine. By inventing such devices, they had this idea that because it is only human intelligence that is capable of performing arithmetic operations, if we can apply this ability into machines, these machines have at least those properties of human intelligence. In their idea, they built a machine that could think instead of them and much faster than them.

The history continues with two important figures. George Boole had the idea that the human thinking is guided by the laws governing mathematics and logic; therefore proposed his Boolean algebra that in his view could be used to investigate the "Laws of Thought". Gottlob Frege, that is considered to be the father of analytic philosophy, was the first mathematician to conceive the idea that mathematics can be reduced to logic. Frege expanded the idea by including language and after inventing his own version of logic, tried to prove that both mathematics and human natural language is based on logic.

The valid conclusion that can be made in that part of the history is that by proving the reduction of mathematics and language in logic, we can not only finish the major philosophical debates in the course of the history of western philosophy, but we will be able to build machines that are guided by these logical rules and can therefore resemble human intelligence.

The history again continues with other two important figures working on two sides with the same assumption mentioned above. On the one side Bertrand Russell and his collaboration

with Alfred North Whitehead resulted in an attempt to reduce mathematics into logic in their seminal work: *Principia Mathematica*. On the other side Ludwig Wittgenstein tried to found the language on the principles of logic in his *Tractatus Logico-Philosophicus*.

Wittgenstein had eventually found that such an attempt was to a large extent futile and had changed his attitudes in his later period. It was Russell's reductive view that came down to Alan Turing as a legacy. Turing proved that he could build a machine that can compute anything that is computable. His idea laid the foundations of today's computers and the assumptions behind Artificial Intelligence.

Computers can do any mathematical calculation and if we look closer, we see that they are doing so by logical gates. That was exactly Russell's idea and somehow the realization of his book, but several points need to be mentioned about this.

I make a distinction between the result and the method. Considering the abacus as the simplest example of a smart machine, although it does the arithmetic calculations and therefore replaces the human mind in doing so, the only resemblance is in the input and output. The method of abacus in calculations is absolutely different from that of human mind. The difference between methods is very important and is nearly ignored throughout the history of Artificial Intelligence.

To give an example, by considering my own method for addition, I do addition much with the help of memorizing. When I see six plus eight, the fourteen immediately appears in my mind. The same for multiplication. Now I ask: does the abacus do the addition of two numbers by memorizing?

This is the case with all of the computing devices; from the abacus and Napier's Bones to modern computers. Concerning modern computers in solving more complex mathematical formulas, another problem raises: we not only use different methods from computers, we do not even know the exact method and difference between human individuals in solving them. The method that X uses in solving an equation may be different from that of Y's. The third problem is that I do not know exactly how I solve an equation. It is I guess to some parts by memory, some parts by picturing it in mind and so forth. So in solving a problem, vision, memory, hearing and many other physical and mental abilities are directed toward the problem.

One might say that the difference between the methods is of no importance. I accept this complaint only if we use the abacus and other devices for practical purposes and no more. But when it comes to call them "smart" machines, the difference in methods is of the utmost

importance. The problem of the theories of Boole's and Frege's is that they thought because they could create formal systems that was to an extent able to cover many aspects of human problem solving and language, they had the idea that the entire human mind can be reduced to their formal systems. The emotion, which is in my view the primary force of human mind, is completely absent from their formal systems.

In other words, using an uncontroversial logical truth which is the Indiscernibility of Identicals, if the machine intelligence is identical to human intelligence, then the two must have the same properties. In this case the method is a property which is not equal in human and machines.

The strategy that human reason devises to solve certain problems with certain methods is never like its own strategy and if we want to call a machine "smart", that machine must follow the same strategy of the human mind. The problem is that we are unable to easily understand the mechanism of the mind in problem solving. The next important question is: can we ever know the mind's method?

Conclusion on AI

In this article I made my arguments over some of the active areas of artificial intelligence. There are other areas which are studied in most of the universities under the AI subject, like Artificial Neural Networks or Evolutionary Computing, which are more focused on specific problem domains and applications in prediction or general problem solving in optimization problems.

Although no one doubts the contributions of AI scientists and researchers to useful technological advancements in many fields like medicine, these contributions are not related to any progress in the implementation of human mental power into computers. For an objective viewer with a cursory glance, areas that define today's AI are working together to bring about the grandiose promises of early pioneers of the field. But with further research the viewer will find out that these areas, however useful, have applications in the world of technology and have no connection with the implementation of human mind and also many of their discoveries are not even related to each other.

As long as human intelligence, even with regard to diverse definitions of intelligence, is relevant to human being, is a special manifestation of emotion and in every human activity we can see its presence and imprints. In my view, the possibility of AI can be summarized in this

question: do we have the ability to know enough about the essence of emotion to implement it by the machines?

I did not go through the details of whether we can ever have sufficient knowledge about the emotion. Thinkers and researchers of other fields like epistemology and psychology have tried to answer this question and it seems to be a perennial problem. I will discuss my opinion about the essence of emotion later in this article.

Cognitive Science

Cognitive Science, as its name implies, is a scientific field that incorporates several other scientific disciplines like psychology, neuroscience, anthropology, linguistics, artificial intelligence, and any other field that tries to know the human mind. The reason of such an interdisciplinary approach is that it can get insight from the discoveries of these fields and therefore present more satisfactory explanations of its theories about the mind. It follows an information-based idea and focuses on how information is represented, processed, and transformed within human mind and consequently model it with computer to assess the validity of its hypotheses. To summarize the fundamental concept of cognitive science, I refer to the statement of Thagard: "thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures." [2]

If one is familiar with the fundamental tenets of the Artificial Intelligence, he recognizes the close analogy between the assumptions of AI and those of cognitive science. I put the simple assumption behind both fields in a nutshell as: the brain is a central processing unit which is composed of neurons; our understanding of mind and thought is similar to programs that are running on this machine.

So, the obvious conclusion is that both fields can be reviewed if we bring the shared assumption under scrutiny. Above I presented my arguments over AI that covers many aspects of cognitive science. Here I continue with another aspect which is the key point of the assumption: thinking as computation.

I put my argument in brief. Human mind and all of its properties are absolutely under the influence of entities, like emotion; therefore because we can never bring, for example, emotion into the computer hardware and software, we will never have the genuine mind or intelligence into computers.

Most of the efforts of cognitive scientists have focused on how to consider the mind as a set of computational processes that are governed by certain algorithms. Computer modeling is the term that can be found in most of the papers in cognitive science journals. Scientists in the field claim that they can model mental characteristics like perception, memory, learning, problem solving, creativity, and nearly everything that is considered to be mental process. Some of them even claim that they modeled human emotion which in my opinion is among the most ludicrous claims. I do not doubt the usefulness of modeling in other scientific fields like meteorology, but in knowing the mind computer simulations and modeling are futile.

Improper terminology like "mental process", "representational structures of the mind", and "information representation in mind" that is so prevalent among most cognitive scientists indicates that they convince themselves of the reality of their erroneous analogy between computers and human mind.

I think the fact that emotion, which I argue to be the background force behind the mind and generally human existence, has been ignored or de-emphasized by cognitive scientists is because they have recognized that their computational approach is unable to explain emotion. By taking a glance at the scope that cognitive science covers, we can find perception, action, decision making, memory, learning, language processing, attention, and everything that in their view is not related to emotion. But as I tried to explain in the previous sections, it is impossible to separate the emotional background from these items in human mind.

Unlike cognitive science, other disciplines that contribute to cognitive science have not been useless in achieving their own goals. For example, neuroscience has been successful in treating people who suffer from many brain dysfunctions or injuries.

One might ask that if I am right in my argument against cognitive science, there should be no discoveries and findings in the sixty-year history of this field. First let me define the goal of the cognitive science: understanding the mind and its cognitive processes that leads us to have a clear causal relationship of different components of the mind. This understanding makes us able to discover every aspect of mental function like the prediction of a choice in decision making. In brief, the field aspires to define unalterable rules that govern human mind and behaviors, inspired by discovering the rules in natural sciences like physics. If we consider the goal of the field in this way, cognitive science is not even close to its fulfillment. It is only by changing the goal to other fields that we see findings. For example the introduction of computers into neurology helped brain scientists to uncover the root causes and results of

specific brain dysfunctions. That is regarded as one of the findings of cognitive science, but is such a finding in accordance with knowing the mind that is the ultimate goal of cognitive science?

Affective Computing

Having formed the ideas that I have mentioned so far, I unexpectedly became aware of a discipline which, although with a different method, has discussed the same issues. It was Affective Computing. The major difference is that it is a scientific field and accepts only scientific discourses. Bellow I will introduce it and discuss important papers and publications concerning those parts that I am interested. At the end of every part, I will write my personal opinion about it.

Picard's 1995 Paper

Rosalind W. Picard's 1995 paper that was presented as an MIT Media Lab. Technical Report was probably the first essay that presented and unified different aspects of relationship between emotion, intelligence, and computer. [3] She has made a respectable contribution in attracting attentions to the primacy of emotion and its role in human intelligence. I will outline her historical paper in which she coined the term "Affective Computing".

In her paper, she first recognized the importance of emotion in human intelligence and decision making. It was so interesting for me to see that she came into the same results that I came; but with exactly the opposite method. I used the introspection and philosophical arguments but she used scientific discoveries of neurologists. In brief, neurological evidences emphasized the importance of limbic system in the brain which is the seat of emotion, memory, and attention. If anyone does not accept my introspective claims over the primacy of emotion in most aspects of human intelligence, I am glad that there are scientific researches buttressing my ideas and I recommend these people to read Picard's paper and its relevant references to the neurological literature.

Although she recognized the primal role of emotion in her entire paper, after the first part, she swerved into a relevant but application-oriented discussion over the "recognition" and "expression" of emotions by the computer which is in my opinion an entirely distinct issue. She proposed models and methods like HMM for this end. In fact, to put it crudely, she misused the importance of the first part of her paper for a different but applicable field of

mimicking and recognizing emotions in human which is more relevant to statistical pattern recognition.

In the third part she went to count the possible hazards if the idea of emotional computers was realized. Again, I distinguish between the ability of expression and truly having emotions; because the possible dangers would be different. But I found the discussion over this to be pointless as long as we have not gotten close to either of these ideas. It is like counting the possible consequences of a real artificial intelligence without having any idea whether the dream of an intelligent machine is possible or not.

The last part of her essay counts the applications of Affective Computing only in terms of "recognition" and "expression" of emotions. There are indeed applications from entertainments and films to aesthetic tastes.

Although I express my gratitude for her contribution to Affective Computing and establishing it as a field that considers the emotion as the center of its activities, I criticize this paper for not being able to discern the important differences between "recognition" or "expression" and truly "having" emotions. She writes in the summary of this paper: "… to build computers that make intelligent decisions may require building computers that have emotions." But she only focused on the proposed models and applications of machines that recognize and express emotions. In my view, she introduced the Affective Computing as having two distinct sides; the side of recognition and expression of emotions which she preferred to discuss more, and the side about computers that "have" emotions that she only mentioned in her essay. I believe it to be the major and more important facet of Affective Computing because even if my introspective arguments and beliefs are wrong and we accept only scientific findings, we will admit that without the ability to apply emotions into computers, we will never transcend a limit in computer intelligence; therefore we ought to focus on possible methods for this idea. Even if we find it impossible to bring emotions into computers, proving the impossibility would be of great importance to not only artificial intelligence and cognitive science but many other scientific and non-scientific disciplines.

Picard's 1997 book "Affective Computing"

This book can be considered to be an important step toward introducing the Affective Computing as an established discipline. [4] The structure and distribution of subjects is more or less similar to her previous paper, but on every subject, indeed with more details. As I

indicated above, I will not discuss the majority of the subjects which is about recognition and expression of emotions, potential applications, moral and ethical dilemmas, low level implementations like signals and systems, wearable computers, and many other closely related issues that are mentioned in the book. Instead I will focus on the part that is about computers that have emotions and her mindset about this topic.

The first relevant part is in second chapter on "Computers that "Have" Emotions". In brief she posed the question "What does it mean, computationally, for a computer to "have" emotions?" Then she proposed that five components would suffice for any system to have emotions. Most of her arguments about this topic revolved around these five components. I will outline these components in short.

Component 1: Emergent Emotions and Emotional Behavior

This is about every system that has convincible emotional behavior and related outward expressions. She writes: "Simply expressing an emotion can be seen as a kind of emotional behavior."

Component 2: Fast Primary Emotions

This includes those involuntary motor or expressive emotional responses that are innate in humans. These behaviors happen before we become aware of the situation and mostly in response to a threat.

Component 3: Cognitively Generated Emotions

Every system that has the ability to generate or alter emotions by explicit reasoning has this component. It happens, for example, when we try to assess our emotional state or alter it by deliberate thoughts. "If merely thinking, "Don't worry; be happy" causes you to be happy, then that happiness is cognitively generated."

Component 4: Emotional Experience

This component is composed of three aspects:

-Cognitive awareness: every system that is cognitively aware of its emotional state.

-Physiological awareness: every system that is aware of its physiological changes that are the result of an emotional experience.

-Subjective feelings: the most familiar aspect of emotional experience which is the personal feelings or "gut feelings".

Component 5: Body-Mind Interactions

This component emphasizes the importance of emotions in visible or non-visible physiological changes and influences on other mental activities like memory. This fifth component is a general description of the influences of emotion in humans.

Before concluding the section on components, she went on to add that computers would not need to have all of these components and it is only human that has all of these in one system. She writes: "… not having all five does not imply that the system does not have emotions" and she gives examples like patients who cannot communicate their emotions, although they have them or patients who suffer from impairment in emotional experience that cannot make decisions but still have emotional behaviors and so forth.

Then she writes about possible tests to evaluate the performance of the system that has each of these components. She counts five separate tests for each of these components.

After that she discusses computers with minds and bodies, possible survival mechanisms, self-awareness of robots about their physical states with sensors, and some possibilities about implementing positive and negative feelings.

The next part that is relevant to our discourse is at the end of the second chapter, named "A Note on Imitating vs. Duplicating". This is, in my view, the most important part since it is about whether these activities will lead to the systems that actually have real emotions or just a simulation of them. She answers this question by giving examples of recognition like that of computer vision. She argues, in case of vision, "… although we are a long way from duplicating the human visual system in computers, nevertheless, at certain levels a computer can still be said to "have vision abilities"." She has the same opinion about other issues like having emotions. She writes: "… there is reason to believe, allowing for a kind of computer architecture that not only contains multiple processors and communication pathways, but also multiple sensors and regulatory mechanisms, that a computer could be built that implements mechanisms like those in human emotion." She believes that at present time, means of implementing the fourth component is a mystery; however if it can be implemented then we have the prospect of duplicating the inputs, outputs, and intermediate influences of a human emotional system.

Later in chapter seven, she proposes some recommendations for how to implement the five components of emotions. She begins with the third component and introduces the OCC model. In short this model is a rule-based system in which input of each rule is the desirability of an event and the output of it is the generated feeling. Later she discusses this model in detail

for other uses and components of emotional systems. After that she introduces another rule-based model named Roseman Model which can be used for human appraisals of situations and how attention in different things can generate different emotions.

After the two models which are mostly relevant to cognitively generated emotions, she names three models that can be used for non-cognitive generation of emotions. The first is the Cathesix model that focuses on four elicitors of emotions in human. These are neural, sensorimotor, motivational, and cognitive. The mechanisms in this model are not much different from OCC and it only differs in parts like updating rules. The next model is three-layer architecture. These layers are reactive, deliberative, and self-monitoring. In response to certain stimulus, one of these layers and its rules are activated. This proposal does not have the details of implementation and is only a recommendation of different models for different layers. The last model is about simulating physiological or bodily changes induced by emotions that can happen in computers. It includes varying temperature, degrees of perception of pain, and so forth.

The next part is about models for implementing the influence of emotions in agents. The computer program ACRES tries to learn about emotion from a user and then put its learning in practice. "In general, ACRES diagnoses the situation over time, generates an emotion, and chooses meaningful action."

Later she discusses and proposes possible methods for implementing the influence of emotions on learning, decision making, and memory.

At the end of the seventh chapter like that of the second, she discusses important issues. She argues that even if we succeed in bringing accompanying physiological and hormonal changes in computers, their experience of emotion will be different from ours. She believes that the difference in emotions is because of the difference in physical aspects of computers and humans. She writes a noticeable paragraph: "… the emotional experience we can give to a computer does not duplicate that of humans; computers cannot feel what we feel. But for that matter we cannot verify that our own children can feel what we feel; we only guess that our similar physiology permits similar experiences."

In the last part, namely "Emotion as an Umbrella", she argues that the emotion should not be treated as a single concept, but composed of many different components with different models and mechanisms for each. She discusses that for example, low-level components require signal-based representations and connectionist interactions; whereas for high-level

emotions, rule-based reasoning can be used and it is not appropriate to use one rule-based system for all emotions. In the end we can combine these different mechanisms and their relationships in one system and we should ensure that they are working cooperatively.

That was an abstract of the relevant parts of her book about computers that have emotions. By pondering over my given abstract of the book, I guess we can have an overview of the mindset of the author about this issue. I will give my own opinion of her attitude in the following.

The problem begins from the beginning. She counts five components of every system that can have emotions. The basic problem, which is a bold claim, is that systems, in technical terms, cannot have emotions. This is another one of those anthropomorphisms which is based on the naïve idea of presupposing the existence of true emotions in machines. Picard's question can be summarized as this: how do you know if anything has emotions? Most of those five components are about how to build a system that can fool the observer of having emotions. She, deliberately or not, does not make the simple distinction between fooling the observer and having emotions. Of course that is possible to build a machine that, not only at first sight but even with spending some time, can make us believe that it has emotions; but does that machine have real emotions? That seems a very simple question but I am really at a loss why she, as a professor at MIT, does not care much about it, at least in this book.

Someone might ask about my criteria for anything to have emotion; my answer is: every cellular and carbon-based living organism that struggles for survival and reproduction has emotions. That is my simplest and the most straightforward test for this purpose. Everything else either tries to fool us or is our own tendency to anthropomorphize.

Patients who suffer from expression of emotions have emotions indeed; just like those who cannot make decisions. Even for patients with impaired limbic functioning, there is no doubt that these humans have emotions; although their impairment may lead to indifference in survival instincts that can distort their emotional experiences. If they were living in wild nature, this indifference would cause them death. That is the result of impaired-developed emotions.

At the end of the second chapter in "A Note on Imitating vs. Duplicating" she gets close to the truth but she evades it. She believes that by working more on technical aspects like vision and computer architectures, we will duplicate human emotion-induced performance in machines. I wish she had defined the meaning of duplication more clearly. If by that she means

to try to imitate only human performance, then why bother to discuss emotions. I guess she had lost the point. Like her previous paper, at first she truly mentions human emotion's influence on human performance, but later forgets that she should have focused on implementing human emotions into machines. That is clear that if we implement anything else than human emotion, we should not expect to see the same performance of humans. So what are we looking for?

In chapter seven, the proposed models and architectures are the result of the same catalog of mistakes. There is not even a single new idea in them. They remind me of the rule-based expert systems that were popular in early 1990's. They are representatives of the failed symbol-manipulation approach to AI that only assigns the name of emotions as merely internal states of machines. The ideas like increasing the temperature of the machine to imitate the corresponding physiological changes in human looks more like a subject of comedy.

In the last part of the seventh chapter, she truly writes: "… computers cannot feel what we feel." I believe nothing can feel what we feel. Our emotions are subjective and I came to believe in the influence of emotions in my life only by introspection. It is my subjective feelings that tell me that flesh-and-bone made people around me have more or less similar emotional capabilities of me. Now one may ask that isn't it possible that I have been fooled in my whole life that these people around me do not have emotions? We can continue this reasonable skepticism on every other subject, but again I listen to what my instincts tells me which is to believe that these people, like me, have real emotions.

In the last part she, like many contemporary logicians, makes an attempt to approach the implementation of emotion in an analytic way. In her view different emotions should be put into the machine using different mechanisms; some with higher levels of abstraction, like rules, and some with lower level hardware implementations. This approach presupposes that we have a number of distinct emotions and each can be treated and implemented separately. Many similar attempts to list a number of emotions have failed, not because of the complexity of the problem at hand but because such an approach is a very naïve and simple-minded.

I do not want to go through the details but I think I have to make it clear why I call that approach naïve. I claim that there is no objective measure or definition for specific emotions. As a child, we learn to assign words to our natural responses to certain environmental stimuli. We then become certain that if other people say that someone is angry, he is experiencing the same emotions like when we were angry. We involuntarily empathize; although we have no evidence that that person is going through the same feelings as ours. The only evidence is our

personal experience. That is the reason why I believe that true meaning of anger is subjective and personal. Now how to define anger objectively? Some may say that there are shared visible measures for us to recognize someone as angry. That is right but I ask: is anger those physical and visible traits or a kind of personal and subjective feeling?

Another criticism to such an approach is that I believe that emotions and their generation are from a holistic and inseparable entity. When our emotion changes from anger to sadness, there is no single unit that stops working for another unit to start. The distinction between different emotions is evidently doomed to failure.

Of five components that she mentions, only the third aspect of the fourth component, which is subjective feeling, is closer to the truth. She recognizes the importance of this, but it does not seem that she had spent enough time to think about it and she treats this one like other components. She only mentions that if this component can be known and implemented and would no longer be a mystery, we will succeed in our goal. I believe by ruminating on this single aspect, we will no longer consider other components and possibilities.

Other Important Relevant Papers

There are two other papers by Picard in 2003. The first is "Affective computing: challenges" printed in the International Journal of Human-Computer Studies. [5] In this paper, she does not discuss the subject of computers that have emotions; instead she responds to criticisms and challenges about recognizing, expressing and modeling emotions.

The other paper is "What does it mean for a computer to "have" emotions?" [6] The structure and topics discussed in this paper is to a large extent like the subjects that I separated and outlined from her 1997 book. One difference, for example, is that she omitted the third component, namely *Cognitively Generated Emotions*; therefore the remaining is four components of a system which is considered to have emotions. But the most important difference is that she confessed several times that having real emotions in machines is a serious challenge. Although she repeats the same mistakes as she did in her book, I recognized a more realistic approach, especially about the component "Emotional Experience". I will briefly discuss important parts of this paper.

She counts four applications of affective computing which are: building humanoid robot, intelligent machines, understanding more about human emotions, and making machines less frustrating to interact with. Having recognized the problems associated with giving

computers real emotion, I guess, she writes that she is interested only in the fourth. She argues about how to accomplish such task without having to give computers emotions. She writes: "I have no longing to make a computer into a companion; I am quite content with it as a tool." Later she writes "… we should not let mere complexity fool us into thinking emotions are there." In the component "Emotional experience" she argues that this ability lies in human consciousness and she is not convinced that we can one day understand and implement it; nor is she convinced that we cannot. She writes an important part: "I do not yet see how we could computationally build even an approximation to the quality of emotional experience or experience of self that we have. Thus, I remain a skeptic on whether machines will ever attain consciousness in the same way we humans think of that concept. Consciousness, and life, for that matter, involves qualities that I do not yet see humans as capable of creating, outside of procreation. Perhaps someday we will have such creative abilities; nonetheless, I do not see them arising as a natural progression of past and present computational designs, not even with the advent of quantum computing."

The meaning of the word "duplication" that I complained about in the previous section is mentioned by her, saying: "I use the term "imitate" instead of "duplicate" with respect to implementing this component in machines."

Although she repeats the problematic claims that we do not have to give all of these components to machines to call them emotional, it is obvious that she eventually faced the inherent problems associated with Affective Computing.

Scientific Evidence: *Descartes' Error*

Because of my philosophical studies over the inherent unsolvable problems in induction and scientific method, I personally have never depended solely on scientific discoveries; but for those who only accept scientific findings, I present an important reference supporting my claims for the primacy of emotion. The reference is a book; *Descartes' Error* by Antonio Damasio published in 1994. [7]

This magnificent book, which is written by a cognitive neurologist, is in my view an important book that, although is considering mind-body interaction from a scientific perspective, has had implications on both philosophy and many scientific disciplines. I highly recommend this book to anyone who is interested in mind-body problem or issues concerning evolutionary neurobiology or in general human nature.

The book is composed of three parts. In the first part he mentions evidences of patients with the brain damage who had undergone psychological abnormalities concerning problems in the generation of emotion. Because of the problem in the proper generation of emotions, those patients had developed difficulties for not only lack of a normal emotional life but decision making, planning, attention, memory, and other cognitive abilities.

By giving detailed technical explanations of brain's structure and anatomy in the second part of the book, he uses the evidences from the first part to assemble a hypothesis. He emphasizes the close relationship of brain and body and their connection with mind. He then, argues that emotions are special manifestations of the desires for survival and reproduction and goes on to discuss the neural substrate for emotions. At the end of the second chapter he gives his own hypothesis which he named the "*Somatic-Marker Hypothesis*" that is about influences of emotions in human mental and physical life.

In the third part he tests his hypothesis on patients and non-patients. He even takes a look on several long-standing philosophical problems like the idea of self, consciousness, and subjectivity from a neurobiological perspective.

To be honest, I was surprised to see that the ideas that I had developed by introspection in my own life and studying philosophy, perfectly suited neurobiological findings of Antonio Damasio. In this book he uses neurobiology to give a general view of mind-body interaction and considers the human body as a holistic entity with inseparable parts which is engaged with a highly complex environment that both acts on the environment and is influenced by it. He rejects every attempt to study mind-brain or brain-body functioning separately and without the influence of the other and believes that in order to gain a comprehensive knowledge for the mind, brain, or body, we must consider their influences on each other and on the environment.

I recommend those who defend the goal of artificial intelligence and the perspective of mainstream cognitive science to read this book; since even if they flatly reject my argumentations, I think, and hope, that they would stop, or at least think about, their notion. At the end of the book, Damasio mentions Cartesian dualism as the source of a wrong perspective that has had pernicious influence not only on the philosophy, but on many disciplines of science and technology. To mention the part which is relevant to our discourse, he dismisses the analogy between mind-body and software-hardware. He writes: "My concern, as you have seen, is for both the dualist notion with which Descartes split the mind from brain and body and for the modern variants of this notion: the idea, for instance, that mind and brain are

related, but only in the sense that the mind is the software program run in a piece of computer hardware called brain". To put his general view of the issue, he writes: "… the comprehensive understanding of the human mind requires an organismic perspective; that not only must the mind move from a nonphysical cogitum to the realm of biological tissue, but it must also be related to a whole organism possessed of integrated body proper and brain and fully interactive with physical and social environment."

Conclusion

From the arguments that I have presented so far, everything which is relevant to human intelligence has to be defined by its relations with emotions. Now the question is what is the essence of human emotion?

First, I have to make it clear that I see the definition of mind as an entity or stuff outside the space and time insufficient. I do not eliminate the mind like eliminative materialists, since mind and mental states are and will always be the best way to describe our thoughts and emotions; but that does not indicate their existence in another dimension or with a mysterious and unimaginable matter. I believe that human knowledge can only reach as far as to know the human being as a physical organism. If there is existence in any other dimension or with a different matter from what we can know, that is outside of human understanding.

The tendency for survival and reproduction, which is indeed physically innate, determines various emotions in us; so the potential for generation of emotions and also its experience is physical. This tendency is shared between all of the animals. In the course of evolution of human organism, the neocortical parts of the brain which is the seat of reason and thinking, has helped us to survive better in the environment. In other words, the brain in mammals evolved as an involving factor between environmental stimuli and proper behavior. To what extent the produced behavior of the brain is conducive to the organism's survival and reproduction determined his fitness in that environment.

So the physical structure of the organism and its behavioral properties correspond to each other. The factors that has influenced each individual to be with specific physical structure that shows certain behavioral attributes, are not out of these two: genetic and environment.

To put it in short, the brain and body should be viewed as two entities but with close and inseparable relationship that can be called human organism. The relationship of this

organism with the highly complex and dynamic environment is for survival and reproduction and this tendency represents itself by emotional experiences that we feel in our lives. It is this organism and its relationship with environment that form our existence in this world. I believe that every human science should define its relationship with this picture: human organism with tendency of survival and reproduction living in a highly complex and dynamic environment.

Behavioral, emotional, and mental differences between individuals are the result of differences between physical structures of their bodily organisms. This difference is not just the difference between brain structures but both brain and body are involved. The structure of this organism is, to some extent, flexible by being constantly influenced by the environment. The degree and intensity of this flexibility is also different between individuals.

With this perspective, there are certain physical rules that govern our brains. I can see the human mind as a function, similar to mathematical functions. I would call it a biological function. Our reactions to choices and stimuli are without a doubt predictable. The problem is that it is too much complex to predict a person's behavior. At first it requires tremendous knowledge of the brain and its structure and relation with body. Then we should have an exact and detailed state of the neurons and the chemical balance of the brain and body. The science has not reached to that point.

The mental states like imagining, thinking, reasoning, memory, and all other faculties which are traditionally considered mental can be defined by the biological structure of the brain. These faculties of the mind that I call physical states are all working together to keep us alive and help us in the perpetuation of our race. That is why they are constantly under the influence of emotion; the emotion which works like a representative of human basic desires and instincts.

I believe that study of the mind should, with the help of disciplines like anthropology and genetics, be restricted to biology and neuroscience.

The important point in this perspective that I have depicted is that the idea of thinking as computation and representation is no longer acceptable. Human intelligence is defined in its biological organism which is composed of its physiology and its relationship with environment. This flexible intelligence can never be brought outside the organism and be implemented into any machine. If there is a machine that has the same properties of human intelligence, it is a living machine with carbon-based cells with all of the physical properties of human body that is no longer a machine but a human individual.

I am not against the benefits and achievements that artificial intelligence has offered in that last sixty years. I only complained about the notion of analogy between human genuine intelligence and artificial intelligence. I think there are still people who believe that computers will one day have real human intelligence with all of its properties. I believe that with the present silicon-based computers, or even with future architectures like quantum computers, we will never be able to have human intelligence in these machines. The only machine that can have human intelligence is human physical organism. We are at best mimicking some aspects of human performance with different methods from that of human. That defines various attempts of different active fields of artificial intelligence today. I only tried to emphasize on the word *artificial* in artificial intelligence.

References

1. Hume, David (1738). *Treatise of Human Nature* (p. 216)

2. Thagard, Paul. *Mind: An Introduction to Cognitive Science.* MIT Press, (1996)

3. R. W. Picard. "Affective Computing". MIT Technical Report #321 (1995)

4. R.W. Picard. *Affective Computing*. MIT Press, (1997)

5. R.W. Picard. "Affective Computing: Challenges", Int. Journal of Human-Computer Studies, Vol. 59, pp. 55-64. (July 2003)

6. R. W. Picard. "What does it mean for a computer to "have" emotions?" M.I.T Media Laboratory Technical Report #534 (2003)

7. A. R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Gosset/Putnam Press, (1994)