The "I" and the robot

Cristi Stoica

Abstract. Man is a lost child, trying to find his way. The main purpose of science, technology, ethics, politics, religion, should be to understand and support humans, their freedom and right to happiness. Can science explain us? Can politics and religion free us? Can technology replicate us? I argue that we know very little about life, consciousness, humanity. Hence, no ideology should be above man. If we accept this, we realize that we can accept others as they are, and not as enemies of our way of life. Our evolution continues, and everyone should be free to find and follow their own path, while happily allowing others to follow theirs. This freedom has to be protected by access to information, education, transparency and critical thinking.

Disclaimer. Because this essay is speculative, I strongly advise the reader to use critical thinking and take everything with a grain of salt.

Ode to man

It is amazing how the universe works, as governed by laws which ultimately are simple, yet combined give such complex phenomena as those we observe. What can be more wonderful than this regularity, parsimony, symmetry, beauty?

Perhaps the fact that this matter became able to look at itself and ask foundational questions about the laws, the nature of existence, even about its inner nature. This is by far more amazing and perhaps more difficult to understand than anything.

I think that living beings, in particular humans, are the most interesting and important things in the world. If there are ultimate physical laws which wait to be discovered by us, so be it, they will always be there to be found. But we can't say the same about humans, this fragile species which was able to rise and try to understand the universe. So one should try and improve as much as we can their brief existence in this world.

Undefining the man

Often, ideologies trying to build an utopian world for mankind, failed really badly. When people didn't care about the ideals promoted by an ideology, they were considered enemies of the good intentions of that ideology, and were repressed. Ideologies fail because are based on *idealization of man*, a simplified model that is supposed to work, like a bed of Procrustes.

No matter how humans differ from animals, they became what they are by change. So no definition should enchain man. The definition doesn't really matter, one should always let humans to be what they want. One should not protect them from evolving, and evolution may mean to take risks and fail many times. Actually, failure seems to be the way we learn.

Freedom to choose one's own destiny is more important than protection against failure.

Does science explain you?

How does science explain what is life, men, and humanity? Is life just self-organized matter? Is consciousness just self-aware matter? Are humans the only conscious living beings?

Visualize one of the persons you love most: life partner, children, parents, *etc.* Now zoom out: see how that picture become smaller, until you see the entire Earth as a blue marble. Continue until the solar system becomes just a dot, and the image can contain our galaxy. Where is your loved one? Thinking how small we are in comparison to the universe, how brief our existence is in comparison to geologic era, we may wonder what's the point of our existence. Now zoom in, see that human's internal organs, the tissues, see the organic molecules composing them, the atoms, the particles. If there is something more than the particles, where is this thing? Where resides the soul of a human? What is life, what is consciousness?

When we try to find the place of humans in our material universe, the entire humanity looks like a speck of dust, like a complex construction made eventually out of particles.

Now close your eyes. What are you? Are you that flesh, that matter? If you replace your arm with a prosthetic arm, will you still be you? What if you replace eventually all of your organs with prosthetics or transplant organs? Perhaps your brain should remain the same, so that you still be you. But how about replacing a neuron with a tiny circuit, performing or not the same operations? How about replacing each of the neurons with some artificial neurons? Will you still be you? Will you still feel that you are you, the one that remains the same while the world around you changes, the one that remains the same while your atoms are replaced?

What are life, consciousness, the "I" that has these experiences, what is to be human? Does science predict or explain these?

The domain of science

There is a spread view that it doesn't make sense to discuss things that can't be checked by experiments, and that they don't even exist. This comes from a particular interpretation of the *scientific method*. The scientific method works by proposing hypotheses, deriving consequences that can be checked by experiments, and performing experimental tests. If the results are not as predicted, they reject the hypothesis. Otherwise, experiments can't prove the hypothesis, they only can corroborate it. Karl Popper described this method [1], which is the best we can do if we want to find a theoretical description of phenomena. He also adds that we can obtain our hypotheses by any means. They may be random or educated guesses, pure creations of the human mind, whatever works. As an extreme example, Ramanujan even claim that his mathematical results were communicated by a Goddess. But no matter what the source is, the hypotheses have to be questioned and tested[1].

When it comes about scientific research, and communicating the ideas to others, one can rely only on the experiment, and on logic. However, there is no reason why the universe is such that we can verify any truth. The idea that humans are able to verify any truth about the world is quite supernatural, because it is very unlikely that a subsystem (humans) is able to extract any information from the whole system (the universe).

Artificial intelligence

> I've seen things you people wouldn't believe [...] All those... moments... will be lost in time, like... tears... in... rain.

<div align="right">

Roy Batty, replicant [2]

</div>

However, let's go along with those who believe that nothing exists, except what you can confirm by experiments. This view leads straight to the *strong artificial intelligence* (AI) claim, that if a program can answer the questions in a way a man would do, then it is intelligent in the same way as humans are. This is the Turing test [2]. But can such a program feel that it is the one having those experiences? Perhaps one can build some artificial brain running the program, and fire some of its artificial neurons in the same regions that are usually fired in a human brain, when the human feels "I am". Seeing the activation of that brain region in the lab would prove that the machine feels like the human. But is this so? Is this self-awareness?

The strong AI view would claim that only what can be checked from the outside matters. If walks, swims, and quacks like a human intelligence, then it is a human intelligence. If two things present the same observable effects, then they should be the same, because there is no visible difference. Should we care about differences that can't be observed?

Is man an illusion?

Suppose that a human intelligence can be equated with a computer program that behaves identically. The computer program can be written on a compact disk, or printed on paper. Should we run it on a computer to make it "alive"? If we think we should, then we agree that the human observer interacting with the computer gives life to the artificial intelligence. So we admit that there is something special with the human consciousness, the "I". If we think that the program is the same, whether we run it on a computer, or just keep it on a physical support, then we are much in the spirit of the strong AI idea.

A computer program is the same, no matter on what support is recorded. Similarly, they would say, a human intelligence should remain the same, no matter whether it is stored on a human brain, or in a computer.

From skepticism to fantastic

Now, think at computer games involving fantastic worlds, populated with unicorns, deities, mutants with superpowers *etc.* If in the future, we will endow the game characters with artificial intelligence, then they become as real as humans are from the viewpoint of strong AI. We don't even have to write the software, since the specifications are enough.

Thinking this one step further, we realize that any fantastic creatures from legends and myths can be used as specifications for AI programs. Just take the Bible as a specification, and you have a detailed description of God, which can be implemented in a computer simulation. But since the technical specifications and the computer program contain the same information, we can conclude that mythical creatures and Gods from religions exist in the same way as strong AI says that we exist.

Ironically, the viewpoint on human nature which is presented as being the most skeptical, by rejecting what can't be observed – the "I", becomes the most fantastic.

Who is experiencing the illusion?

If human consciousness is just a software running on a flesh hardware, then we can very well be a simulation in a simulated world.

In Western philosophy, the idea that we live in a simulated world is usually tracked back to Descartes, or earlier, to St. Augustinus. Descartes started by asking himself whether everything he thought he knew may be just a simulation (an illusion created by a demon). From his doubt, he deduced that he thinks, and from this, "I think, therefore I am" ("cogito ergo sum").

In Eastern philosophy, the idea that we live in a simulation (*Maya*), and the only real thing is the *self*, appeared thousands of years before, being a fundamental element in Hinduism, Buddhism, Taoism. Meditation and awareness techniques were developed, as means of liberation of the illusion. The identification with our body, with our experiences, with the roles we play in each situation, are viewed as hiding the true nature of our self.

Although the meditation techniques often came with a religious motivation, many sages distilled them from religion, which they considered part of the illusion. In modern times, notable in this respect was Ramana Maharshi [3].

In modern psychology, some meditation techniques were recycled as methods of *relaxation* and *mindfulness*, to reduce stress, anxiety, and depression [4].

For a subjective science

There is nothing in science that could prevent us to build automata that do what we do. If these automata don't feel what we feel, at least they can do what we do when we feel what we feel. Maybe someday one can replicate a person, so that observers chatting with it don't distinguish the copy from the original. But you are inside yourself, so you know better than what an outside observer knows. You know that you are. Can science explain this? Perhaps not yet, but definitely should at least study this. Perhaps there should be a subjective science. Objective science could be checked by outside observers. Subjective science would study that interior activity that can't be verified by outside observers. Feelings like anger, love, happiness *etc*, seem to be observable from the outside, given sufficiently advanced brain monitoring devices. But the most intimate part, the one that tells you that you are experiencing those feelings that are visible from the outside, should be researched by a science of the subjective.

The subjective science would only be able to propose models of the inner experiences, and describe them so that others can try to experience them in their own subjective universe. Subjective science may find that there is such an inner "I", or that there is not, and we only think that we are "I", because we are tricked to think this for reasons related to evolution.

You are a principle

But no matter whether science can explain life, the "I", what makes us humans *etc*, truth is that these things exist. So if we can't infer them as theorems from the current science, let's take them as axioms, and see what we can build on them.

The problem of explaining what are we, what makes us conscious and humans, is very interesting. But we lack such an explanation. So how can we help humanity, when we don't know what humanity is?

Well, we can't wait to gather full understanding. We just have to do something. Just like we eat, without needing to know what biological and chemical processes take place while eating. Until we will have an explanation of what we are, let's just accept our existence as an axiom, and see where this takes us.

The most important thing in the world

If life is what science seems to imply, just self-organized matter, just an emergent foam on the surface of a vast and deep ocean, then what would be so special about us? Why would the universe care about us?

Indeed science sees no value in humans, because science is not about value. But how about religion, doesn't it seem to explain our existence? But looking at those religions claiming that God is much more important than us, and He created us with a purpose, we see that in general they claim that the purpose for which God created us is to worship and serve Him.

I tend rather to think that if God exists, He is the one who serves us. We are only children, or rather babies, as compared to how an omnipotent and omniscient being like God is supposed to be. Who would ask babies to worship him, and give them eternal punishment for disobeying, for acting according to their nature? If God exists, then He is great, we are small, and He doesn't need us to worship Him, but we need any help we can get. The idea that we should serve God is just the projection of our fears, the remnants of what our ancestors had to do to survive when meeting more powerful opponents.

Evolutionary psychology may suggest that we consider life, consciousness, the happiness of humanity to be so important, for selfish reasons. Selfishness can explain even why we act altruistic – to get some advantages. People avoid trusting selfish people, so we have to adapt to appear altruistic. And the best strategy is even to believe that we are altruistic, that we care about others, that we love them. If the word would be like this, then it would be very ugly. But remember that we are discussing a materialistic theory, in which consciousness is just a secondary phenomenon, a foam on the ocean of matter, a simulation made out of molecules. So, if we are just a happening of matter, then why would we expect altruism and love be more fundamental?

No matter what the truth is about this, we consider us important because otherwise we would not care to survive. If someone considers he is the most important, then also his life, consciousness, happiness become important. And since we see that others appear similar to us, and they believe in their own importance, and we need them, then we have to agree that their life, consciousness, happiness are equally important. Hence, we care about mankind.

I would like to propose as axiom that

Axiom 1. *The most important things in the world are life, consciousness, happiness.*

The only reason why these are important, is because we have to feel important enough to take care of ourselves and to survive. This may sound very selfish, not to mention circular, but can one name anything else more important than the existence of a human being, the life, consciousness, and the happiness of a human? Although these things are studied by both science and religion, I tend to think that neither science nor religion explains them. But science, technology, philosophy, politics, religion, all should be in the job of humanity.

What humans need most?

The idea that humanity should drive its way toward a perfect society, what can be more appealing? And yet, such utopian dreams were behind religious wars, Nazism, communism, genocides *etc*.

The origin of any ideology that pursues an utopian dream relies on some assumptions about what people need most. Since people are different, they may feel that they need different things. Ideologists of various utopias often see those not sharing their dreams as being evil. They are afraid that opposition and criticism are obstacles in their way to Utopia. This fear makes them try to be more and more in control, at any costs, so they end up building a dystopian, repressive world.

One may believe that we learn from history, and we will never fall in the same trap... Unfortunately, even a single person hardly learns from her own mistakes. How could we expect that humanity, this huge crowd driven more by the psychology of masses, so easy to manipulate, yet so hard to educate, can learn from its own mistakes?

During the entire history, horrific oppression measures, including genocides, were made in the name of utopian dreams, against parts of the population that were perceived as the main obstacle. People learn particular lessons, that Nazism, antisemitism, communism, racism, bigotry of a particular religion are wrong. But the reason is beyond a particular ideology, a particular kind of racism, a particular religion. To be able to prevent this kind of things from happening again, we have to identify the real cause. Not to propose yet another ideology. Give man a guiding ideology, and you give them something to blame for any problem, and something to hate, so you will always have a tension.

The fundamental need of anyone is freedom. Be free, and allow others be free.

As simple as this may sound, dictators always managed to present the things in such a manner that the very victims of their repressive system appeared to be the bad guys. In the political discourse, freedom is very interpretable. Just present the opponents as the enemy of the people, or country, or God, or the utopian dream you sell to them, just appeal to their most fundamental instincts, and you will be able to raise an army to fight any war you want.

If you are religious, you may be tempted to think that those of different religions are to be blamed for anything, and perhaps atheists even more. If you are atheist, you may think that religious people are to be blamed. Politicians blame the opposite party and so on.

Education without manipulation

For people to be free, they have to be informed, and to understand what's going on, what choices they have, and what are their consequences. Therefore, there is no freedom without education. But by education I don't mean manipulation. This is why education must include critical thinking. We have to introduce critical thinking in schools, from as early as possible. It will indeed weaken the obedience of people, the ease of manipulation, the consumption society, but without it there is no real freedom.

Everyone tries to educate you: politicians, commercials, various organizations, religious or not, everyone. They say that they inform you, that they educate you, but they just want you do buy what they sell. No ideology, no religion, no science or technology can help you be free, if you let others think for yourself. The antidote is critical thinking.

Can religion help? Perhaps not, when it promotes the view that man should be punished for not worshiping God, or when asks you to serve God by empowering a certain organization,

by donations of money and time to convert others. Perhaps yes, if it encourages people to accept the ones of different beliefs, if it helps people understand and respect themselves and others. I mean understand and respect, and not merely tolerate. Definitely no, if it teaches people that those which are different are inferior, or less worthy, or bad, and that they deserve to go to Hell and receive eternal punishments for being different. Ethics is for you to know what is right and wrong, and not to give you more reasons to judge and hate others. When you think that people following other paths are less good, and are under the influence of evil, you care less about them.

Can technology help? Definitely. The Internet allows people know each other and understand each other. Sometimes the discussions become heated, so it helps to promote respect on forums. But due to the Internet, people can read about opposite views, and understand them. People will learn more about one another. Also, people will find information which otherwise would be hidden to them, in order to manipulate them. The Internet allows us to find complementary points of view, arguments we normally ignore, it allows us to see that those that are different, are humans too. Allowing the Internet to be controlled by governments or corporations will take this away from us, and it will allow us to continue to be manipulated according to their agenda.

To avoid being controlled, we have to protect out privacy and freedom, in particular on the Internet. On the other hand, we have to make sure all the time that governments and other powerful groups are transparent, and that the society is open [5].

Wealth and freedom

Technology progressed so much in the last two centuries, that now people are many times more productive than they used to be. It is conceivable that now we can work less to make a living. We have longer weekends, indeed, and apparently shorter work days. Buckminster Fuller even said [6]:

> We must do away with the absolutely specious notion that everybody has to earn a living. It is a fact today that one in ten thousand of us can make a technological breakthrough capable of supporting all the rest. The youth of today are absolutely right in recognizing this nonsense of earning a living. We keep inventing jobs because of this false idea that everybody has to be employed at some kind of drudgery because, according to Malthusian-Darwinian theory, he must justify his right to exist. So we have inspectors of inspectors and people making instruments for inspectors to inspect inspectors. The true business of people should be to go back to school and think about whatever it was they were thinking about before somebody came along and told them they had to earn a living.

With every breakthrough, new reasons that keep us busy at work appear. Most of these reasons are due to the desire to have more things, but also the bare necessities became more expensive. Perhaps we buy more than we need, and allow others fill our time working at their dreams. Why not simply follow our own dreams?

By contrast, in the poor countries, people starve because they don't have jobs. So why not solving both problems at once? People from better developed countries can *downshift*, that is, they can work less, to spend more time with their families, friends, hobbies, or their true passion, like researching physics or writing songs. And for the remaining part of their jobs,

they can outsource to people from less developed countries. This is already a way of living for some. People downshift to buy back their time, their lives. They outsource to poorer people, allowing them to survive and make a living without having to leave their country.

Because of differences in economy, people have to leave their homes and go in countries that offer them better opportunities. On the other hand, people from the richer countries feel sometimes that they are invaded, that their jobs are taken by the immigrants. All these problems can and will be solved when people will understand that there is no need to be slave most of their limited existence just to buy more stuff, when they can live their lives and in the same time allow others to live. This will mean better lives for all.

One global country, or countless more?

Because of the modern means of communications, especially the Internet, the boundaries between countries are reduced continuously. In time, globalization will reduce the differences, and the countries will be merely formalities. Does this mean that we will eventually become a single country? Maybe, or maybe not. There is a reverse effect: people find online groups which they can't find in their real lives, because of geographic distances. So, smaller communities emerge. Rather than becoming a huge community, countless many appear every day, of all sizes, of all views.

In time, people will join more groups having similar interests, and they will ask together for the rights which will fit their interests. At some point it may become natural and even possible to create virtual countries, which are not dependent on the geography. So, on the one hand, more and more international treaties homogenize the laws, and on the other hand, people have more specialized needs. I think this requires more flexible, even customizable laws, which guarantee the individual freedom, as opposed to the dictatorship of the majority which is imposed by the current understanding of democracy. It is hard to imagine how these future societies will be. But we have to be prepared, to be open to the fact that each of us are different, have different needs. If we are not prepared to accept the differences, we will feel that there is chaos, that all kinds of people with all kinds of wrong principles surround us.

We have to embrace change and diversity, to admit it, because it is real, and we can't do anything to avoid it. Just like we admit that we have to breathe in order to live.

Cybertopia

Not far in the future, more and more human-like computer programs will be created, so that we can chat online with them and not be able to tell they are not humans. The idea that we can download our mind in a computer, and that computer will behave just like us, will gain more popularity, as is already accepted by many [7].

While computer science will allow in the close future to simulate human intelligence, how can we read the information stored in the brain? Perhaps we can't see at this time a way to download all humans thoughts. But we only need to collect the observable information. There will be plenty such information, on our social network accounts, in our email and messenger history, in the CCTV records, in our own smartphones, in the phone conversations which are recorded and stored by "them". All this information will become available to us (perhaps for sale, since we already agreed that it belongs to these companies when we created our accounts). So, we can in principle build a simulation of a person, and since most

of her friends know it mostly through the online interactions, by phone, email, messenger conversations, they will certify that indeed is the same person.

One may object that the simulation is just a simulation, and it can never be the original. Sure, but who cares? There already are people who believe that there are no differences, because we can't observe by experiments any difference. By then, there will be more. People will want to survive, will want to save their deceased relatives, so they will choose to believe that technology can offer this kind of immortality. So they will use these simulations, and treat them as being real. It doesn't matter if others believe they are not real.

Now, imagine someone who has the means to survive his own death by downloading his mind in a computer. What has he to lose? Many will just do it, because they are dying, or just because they live dangerously and want to backup their mind. But, if you are legally deceased, you are not able to use your own money and properties. So, why not transferring everything, after death, to your replica? If the law doesn't consider the simulation to be a person, the law can be changed. Because science saids there is no observable difference between the simulation and the original, and because the powerful people will want this to be legal, it will become legal.

In time, just like most of us can afford now a computer or a cell phone, more and more will afford to make savings to be resurrected as a computer program, after death. People will want to stay married with their spouse even after "death do them part". They will transfer the simulation in a beautiful and young-looking robot or clone. But then, those having robots as partners, for sex and conversation, will also want that their partner to be recognized as a person, even to make a family. Why not, since there is no observable difference between a simulation who existed before as a human person, and a robot built from scratch.

Being a cybercitizen will come with some advantages: you will be better at computer games and even at your job, because you will not be slowed down by your limited body. You will be able to live in a machine, and perhaps you will not care much about real estates, because you will only need virtual estates. So, you can afford to use the money people usually spend on a mortgage and a car, to do something better, such as to become faster at your job, to buy yourself better upgrades. This means that living people will not be able to compete cybercitizens at work. What will happen then? Will they feel pressed to renounce at their bodies and migrate to cyberworld?

Will we disappear as biological race, being replaced by simulations? Or they will just replace us at our work, allowing us to be free of the need to work in order to live, and spend our brief existence doing what we really like?

Our legacy

The future is unpredictable and open, and we can make it better, for future us and for our children. We want them to live in peace and happiness. They can't, if we want them to continue our fights and wars against others that are different, or to pay them back bills we inherited from our ancestors. The legacy we leave them should be a healthy planet, good relations with others, access to education, freedom, a healthy and critical way of thinking. We have to learn to be free, and to allow others to be free, because this is the only way our children will be happy and free. Then, they will be able to focus on any problems the future may reserve them.

References

[1] K.R. Popper. *The Logic of Scientific Discovery (Routledge Classics)*. Routledge, 2002.

[2] Alan M. Turing. Computing machinery and intelligence. *Mind*, pages 433–460, 1950.

[3] Paul Brunton. A search in secret India, 1934.

[4] B. Khoury, T. Lecomte, G. Fortin, M. Masse, P. Therien, V. Bouchard, M.-A. Chapleau, K. Paquin, and S.G. Hofmann. Mindfulness-based therapy: A comprehensive meta-analysis. *Clinical Psychology Review*, 33(6):763–771, 2013.

[5] K.R. Popper. *The open society and its enemies*. Princeton University Press (Princeton, NJ), 1963.

[6] Elizabeth Barlow. The New York Magazine Environmental Teach-In. *New York Magazine (30 March 1970)*, page 30, 1970. http://books.google.ro/books?id=cccDAAAAMBAJ – last accessed March 26, 2014.

[7] Ray Kurzweil. *The singularity is near: When humans transcend biology*. Penguin, 2005.

Notes

1 Almost all Ramanujan's results were proven mathematically to be correct, a small number turned out to be false, and some conjectures remained to be proven.

2 "Tears in rain" soliloquy from "Blade Runner", Rutger Hauer's version.