Proceedings of the Towards a Comprehensive Intelligence Test (TCIT): Reconsidering the Turing Test for the 21st Century Symposium,
Aladdin Ayesh, Mark Bishop, Luciano Floridi, and Kevin Warwick (Ed.), at the AISB 2010 convention, 29 March – 1 April 2010, De Montfort University, Leicester, UK

Computer system that likes chess

Pawel Dybala[1], Rafal Rzepka[1], and Kenji Araki[1]

Graduate School of Science and Technology
Hokkaido University
{paweldybala, kabura, araki}@media.eng.hokudai.ac.jp

Abstract. In this paper we present our approach to the old computer science dilemma: is proper behaviour enough for a computer system to be considered intelligent or human-like? We discuss the two most classical approaches to this problem, presented by Turing and Searle, and present our own argument by analyzing a simple situation, possible to happen in a real life. We discuss the consequences of taking such approach as ours, present the whole problem from the eyes of an average user, point out some future directions for the field of HCI and propose how Turing Test could be used to achieve some important goals in nowadays science.

1 INTRODUCTION

In recent years we could see numerous projects aiming at creating human-like, conscious machines. A good example of such a venture is the LIREC Project [1], in which European researchers are working on creating artificial companions able to build long-term relationships with humans.

The exciting idea of constructing such machines is what makes people want watch so many science fiction movies. It is also one of major depressive factors for AI scientists, when they discover that it is impossible to construct such machines, as they only operate on zeroes and ones, and even if we make them work also on threes, fours or even hundreds, it will not change the fact that it is hard to imagine mathematic operations conceiving consciousness. Despite this, many researchers, not only from the field of AI or computer science, but also philosophers or cognitivists, still continue endless discussions on how to construct machines that would think or be conscious.

As a matter of fact, this problem is not restricted only to consciousness of computers. In recent world of science we can see numerous research projects aiming at constructing machines that think, talk, experience emotions, are able of liking or disliking things etc. Also here it is quite easy to fall into depression, when you realize what computers really are and how incapable of doing things like these they will always be. Actually, they cannot even perform such simple actions as adding numbers, which was explicitly showed by Levesque [2]. Thus, how can we even think of making machines that can think?

This makes the whole idea of conscious computers seem not worth fighting for. The very basic concept of AI loses its sense, as one may think that constructing truly intelligent machines will never be possible.

This pessimistic argument is not new, as its various mutations were presented by many scientists over decades. One of the best known ones is the Chinese Room thought experiment, in which John Searle (criticizing Alan Turing's concept) claimed that it is possible for to act as if one knew Chinese, which requires only a book with sets of proper rules for this language [3]. This is obviously similar to what computer systems do, making Searle state that behaving AS IF one was doing something is not equal to actually doing it.

Both Turing's and Searle's arguments have been widely discussed, dividing the world of science. Turing's followers claim that getting the behaviour right is enough for the system considered to be intelligent, while Searle's followers argue that acting as if does not equal doing things.

In this paper we would like to present our approach to this subject. Both Turing's and Searle's arguments base on their thought experiments – however, such experiments have one serious drawback, namely: they are theoretical. One of the main sources of criticism towards these approaches is that it is virtually impossible to actually recreate their settings (e.g. the Chinese room) in the real life. Thus, we decided to illustrate our approach with something much simpler – an analysis of a hypothetical, but very trivial situation, likely to occur in the real life. By this, we show how human-computer interaction can be similar to human-human interaction on a very basic level. We discuss conclusions coming from this fact and point out their practical consequences for the world of AI.

2 THE TURING TEST AND THE CHINESE ROOM

In his work published in 1950, Alan Turing [4] stated that getting a computer system's behaviour (in the long run) right and making it indistinguishable from human should be enough for the system to be considered human-like. This is what the famous Turing Test was designed for – to conduct experiments investigating if human judges are able to tell the difference between computers and humans.

Although more than half a century old, this idea was and still is widely discussed. The most influential argument against it, known as the Chinese Room, was presented by John Searle [3]. He proposed a thought experiment, in which a monolingual English speaker is locked in a room and given a book in English, containing a set of all rules needed to process input and generate a proper output in Chinese. The person in the room is given pieces of paper with Chinese characters, and, using the book, processes it to create an adequate response.

This thought experiment, says Searle, proves that one can imitate being able to speak Chinese, without actually knowing it, using only the set of rules. Thus, getting the behaviour right is not enough for anything to be considered conscious.

The Searle's approach, despite having a huge impact in the world of science, causing even questioning the basic concept of AI, was also criticized throughout the 30 years since it was proposed. Levesque, for instance, [2] argued that in the real life the "magical" book with all necessary rules for Chinese cannot exist. He proposed another simple thought experiment, called the Summation Room, in which he argues that the complexity of any set of rules, allowing imitating human behaviour in the longer run makes it virtually impossible to construct machines that would actually do that, due to the enormously long time it would take to process such amounts of information.

French [5] goes even further in criticizing the Chinese room, saying that the question underlying the argument is wrong in the first place. Instead of asking "What are the implications of the fact that someone is answering questions in Chinese without knowing Chinese", we should ask if the very idea of such situation makes sense at all [5]. Needless to say, to French the answer is no, as he argues that the whole setup of the thought experiment is impossible to recreate in the real life, also stating that no such rule-books can exist in the first place.

As we can see, there is still much doubt concerning these two ideas, and the discussion is still vivid. Thus, with this paper we would like to contribute to this field, showing our approach to Turing's and Searle's conceptions.

3 COMPUTER SYSTEM THAT LIKES CHESS

As mentioned above, thought experiments are too theoretical to be fully trusted. Thus, in this section we are going to analyze a very simple, life-like situation, and next draw some conclusions from what we find out.

Imagine a situation in which two people first meet – let us call them Human A and Human B – and all they do is talk to each other. An example exchange of utterances may then look like this:

Human A: So, do you like chess?
Human B: Oh yeah, I do!

Hearing that, Human A will most probably think something like "OK, if Human B says so, he/she probably does like chess", as, looking only at the utterance, there is no particular reason why Human A would think different. Thus, Human A's knowledge is derived on what Human B said, and a logical consequence of this is to say that to Human A, Human B likes chess.

Now, let us imagine a similar situation, in which Human A is talking to a computer system – let us call it System C. If Human A uses the same approach and asks the same question – "Do you like chess?", and the System C answers in the same way Human B did ("Oh yeah, I do!"), logically speaking, Human A should assume that System C does like chess, as it says so.

Thus, in a very simple way, a computer system that likes chess – or is believed to like chess – can be created.

4 DISCUSSION

Needless to say, above experiment would not work like that in the real life – not today, anyway. The main reason why it would not is that humans are aware of the fact that machines cannot like things, as they are machines, not humans.

This is, however, where the Turing Test should prove useful. If the dialogue between Human A and System C was conducted after hiding the latter's identity, say – occurred on an on-line chat channel, Human A would then have no particular reason[1] to doubt System C's words. Thus, in Human A's reality, System C would like chess.

The problem underlying this approach is not new. The knowledge of the fact that the conversation partner is not human was the main issue of the Turing Test concept. In fact, it can be said that in comparison with humans, computer systems start from a much worse position – when we talk with humans, we do not wonder how they are build, how their brain works or what they have inside. Instead, we just assume that they think, feel and like things, as we know that this is how humans are. Contrary to this, in case of computers we do wander how they do what they do, especially if they act as humans. In fact, we could expect that the more human-like computers act, the more suspicious about their abilities humans would be – as they are computers, so, to common knowledge, they should not behave like humans. Even a human behaving in a very odd way would probably be considered more human-like than a perfectly human-like machine.

Thus, it seems like it is all a matter of the right approach. If we manage to convince people not to doubt if creating human-like computers is possible, and to take only their behaviour into consideration, they most probably would start to believe that their artificial partner in fact are human-like, being able to like or know things like we do.

As a matter of fact, on a very simple level it is already happening. Being able to process only zeroes and ones, computers are believed to know how to add, divide and perform much more complex mathematical operations, just judging on the basis of their output. Word processors use thesauri to suggest us what expressions we should use, making an average user think that the software actually understands what they are writing. Average user does not think how the processor or machine is built or how does it perform its operations. All they care about is a proper output, and the fact that it is obtained using methods different from those we use is not even an issue here.

So, here we face the old, good question: is behaviour everything? Turing, in short, claimed that it is, and Searle – the opposite. Both of them supported their theories with strong arguments – which, however, were of purely philosophical nature. We would like to consider the problem looking from the viewpoint of an average user – the same one that does not wonder how the word processor or calculator is built, as long as the output is right.

We say - if we are aiming at constructing artificial companions for humans, i.e. systems that would be able to perform conversations with us, all we should worry is their behaviour. The very term "human-likeness", so often used in

[1] Apart from obviously limited credibility when interacting with someone only through the Internet

HCI and AI in general, means that systems we are trying to construct, should be "like humans", not necessarily identical to humans. If the system acts as if it liked chess, what is the reason an average user should not think that it actually does?

Needless to say, simply saying "I like chess" is not enough. Both Turing and Searle agree that the behaviour should be proper in the long run and on more than one level. Saying "I like chess" may work for the first impression, but if the interaction continues, the utterance itself becomes insufficient. Levesque [2] gives an example of a conversant shouting "I love the Yankees!", which may appear as a manifestation of actual feelings, but is meaningless if nothing interesting follows it [2]. Thus, a system that could be recognized as liking chess would have to act as if it liked chess also in other ways – by, for example, making it topic of conversation from time to time or displaying knowledge about the discipline. If we are talking about embodied agents (robots, humanoids, virtual agents), they could present their fondness also in other ways, for instance - by wearing T-shirts or caps with chess figures. If such sets of behaviours would be recognized by humans as indication of the system liking chess, what is the reason for them not too believe it?

As mentioned above, the main remaining issue then is the problem of knowledge regarding the partner's identity. We know that computers are different from humans, so even if they behave like us, they cannot be the same. As a matter of fact, this starts to become a vicious circle – computers differ from us, so they must behave different, which in turn would make them appear even more different.

Therefore, what we really need here is a break trough in humans' approach, a change in our way of thinking. This could be achieved by misguiding users and making them interact (in the long run and on multiple levels) with non-human partners hiding their identity from them. If the interaction goes well and users would believe that they interact with humans, the true identity of the partner could be revealed. This, hopefully, could convince at least some humans that if computer systems behave in the right way, they could be treated in a similar way to humans.

These, we believe, should be the new challenges for the Turing Test in the 21st century: to find proper methods to convince average users that computers can behave like humans. What we proposed in the above paragraph is only a concept of what it could look like – it still requires empirical confirmation and, when confirmed, more detailed adjustments, like how long should the interactions be and what they should contain. In short – we need to check if this approach actually works, which is possible, contrary to thought experiments like the Chinese room.

If the approach does work, and we actually find a way to make users think that computers can behave like us, it will mean that in our projects aimed at building virtual companions, we can focus solely on their behaviours.

5 CONCLUSION

One may wonder what the above argument is actually about. Making right behaviour the main issue of HCI is not indeed a new idea. Yet, probably due to arguments such as Searle's Chinese Room, in the world of science we sometimes tend to forget it and lose ourselves in endless discussions like "how to make our systems truly intelligent or possessing knowledge". This even becomes the main topic of various symposia and workshops, such as "IJCAI'97 Workshop: Animated Interface Agents: Making Them Intelligent" or "IJCAI'09 Workshop: Knowledge and Reasoning in Practical Dialogue Systems". In discussions conducted during such events some participants tend to present rather pessimistic approach, saying that making computers intelligent or conscious is by all means impossible, as they are nothing more than very fast calculators. Quite popular saying among AI researchers is: "we do not have to worry how we will feel about it when we construct AI, because it is not going to happen".

If we take into consideration the fact that computers are only about zeroes and ones, it can be true – we will never construct AI in the way we imagine it. This approach can make some researchers resign from even trying and abandon their projects – after all, what is the point in pursuing an impossible goal?

However, if we think of computer systems' behaviour as the main issue, we do not have to be such pessimists. If users start to appreciate systems' acting AS IF and start to recognize them as able to be human-like, they will (probably) be satisfied with their new artificial companions. Thus, what we should focus on in the nearest future is:

1) getting the systems behaviour right – so that in all possible dimensions they would resemble humans
2) convincing the users that the fact that their partner is non-human does not have to mean that they cannot behave like us.

If we succeed and manage to construct computers that would act as if they, say, likes chess, which would be recognized by users as such, it would be quite an achievement. And if we do that, who knows – maybe if we create something that behaves in an intelligent way, we will realize that we start to treat it as if it really were? And then, how different for us it would be from interacting with actually intelligent partners?

REFERENCES

[1] LIREC - LIving with Robots and InteractivE Companions, http://www.lirec.org/

[2] H. J. Levesque. Is it Enough to Get the Behaviour Right? In Proceedings of T21st Joint Conference on Artificial Intelligence (IJCAI-09), 1439-1444, Pasadena, California, USA (2009)

[3] J. Searle, Minds, brains, and programs. *Brain and Behavioral Sciences* 3, 417-457, 1980.

[4] A. Turing, Computing machinery and intelligence. *Mind* 59, 433–460, 1950

[5] R. French, The Chinese Room: Just Say "No!", *Proc. of the 22nd Cog. Sci. Conf.*, 657–662, Philadelphia, 2000.