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

Qualia Turing Test
Designing a Test for the Phenomenal Mind

Alessio Plebe and **Pietro Perconti** [1]

Abstract. The Turing test attempted to establish criteria sufficient to accept technological artifacts on the same level as humans. The criteria should establish if a property, acknowledged as human specific, could be ascertained in a machine as well, by a convincing methodology. The property chosen by Turing was *intelligence*, and the testing methodology was a conversation ability challenge. In the current cognitive science, the property that seems to be deeply human, yet mostly resistant to naturalistic and computational explanations, is consciousness, rather than intelligence. Therefore, we believe that a contemporary challenge, heritage of that launched by Turing, could be testing for *qualia* experience in a machine. In this work we outline a sketch of how such test could be designed.

1 Introduction

The Turing test attempted to establish criteria sufficient to accept technological artifacts on the same level as humans. One point to underline, however, is that the criteria is not proposed in order to solve an already standing debate, and it is not conceived as an unbiased judge. The debate is in fact opened by the proposal of the test, and Turing clearly envisaged a criteria such that, at least in principle, machines could effectively meet. He overtly defended this possibility. Now, a good strategy for Turing in defining its criteria, was to first identify a property that at the same time might be ascribed to both humans and machines, but is very human. It should be so human-specific that everybody would agree that machine are quite like humans, if ascertained in them as well. Turing identified *intelligence* as the best candidate property. The great appeal of the Turing test is that it superseded core discussions about what intelligence is, escaping the risk that the inclusion or not of machines would be just a consequence of the definition of the term. It suffices that the overt behavior of the machine in a conversation task is enough similar to that of a human being. In fact one can easily note that in this way, the very property chosen to be the key feature of both humans and machine could be called language. Upon this interpretation, the Turing test adheres to a long tradition, that takes language as the most distinctive feature of humans. He has been, after all, a student of Wittgenstein. Today, on one side there is less consensus that intelligence or language are so special, current biology tends to emphasize commonalities between our and other animal species [7]. On the other side, the common feeling that machines are so different from us persists, it is probably grounded on other properties we possess, that are not necessary intelligence nor language, possibly common to other species, that make us so far from machines. Therefore, if somebody aims to engage herself on the same challenge that Turing started, to place technological artifacts on the same level as humans, she has to search first for an appropriate feature. A feature that, again, if found on a machine, would convince that that machine is wonderfully human. Given such fundamental property, next point is to imagine a test, that possibly, in the same vein as the Turing test, could be performed just by observation of the behavior of the machine. A welcome benefit of a new test, could be to act not just as a dicotomic decision: the tested machine is/is not like a human, but allow for a graduation. Did it happen for the Turing test, in other words, did people become more open to the possibility that machines share something with humans, as long as the performances of automatic answering software progressed? Or did the Turing test remained mainly a philosophical mental experiment?

[1] Department of Cognitive Science, University of Messina, Via Concezione, 6-8 98121 Messina, Italia; Tel. +39 090 47088, Fax +39 090 5728874 {aplebe,pperconti}@unime.it

2 The proposed *qualia* test

Turing's idea had undoubtedly many repercussions in the development of natural language processing, and specifically in automatic answering software, and systems that chat with people on the Web. Although advances in these fields have an important market, we doubt that people are getting more and more convinced that machines are similar to humans, because of its performances. Get used to the fact that machines can reach level of intelligence, with several meanings of the term, similar to humans, people tend to exacerbate what seems to remain drastically different, the phenomenal side, in claiming the machine could never be like humans. In this respect common sense rejoinders with the last trends in cognitive science. In the current debate, the property that seems to be deeply human, yet mostly resistant to naturalistic and computational explanations, is consciousness, rather than intelligence. Therefore, we believe that a contemporary challenge, heritage of that launched by Turing, could be testing for *qualia* experience in a machine.

Note that Turing, with his farsightedness, already discussed the issue of consciousness in his seminal paper [17, §6.4]. His position was that the conversational structure of his test could escape the problem, by limiting the analysis on the observable behaviors, however he acknowledged that consciousness was an unsolved hard problem. Nevertheless, he concluded that consciousness was not central in the debate about machine intelligence, therefore not a strong issue for his test. Consciousness itself has been for a long time the weapon used by many to attack the Turing test in its philosophical foundations [15, §4.3].

On the contrary, we propose that a test derived by the classical Turing test, could be used to investigate if machines can share a phenomenal mind similar to that of humans, in the light of the progress made in philosophy of mind on this argument. The proposal is not

unbiased, as it was that of Turing, in that we believe that the possibility of ascribing phenomenal contents to an artifact is cannot be ruled out in principle.

The question of whether a machine can ever become conscious is not new at all, several works have addressed this issue in the past. A basic strategy in most approaches has been the development of systems that implements some conceptual architecture believed to support consciousness. The architecture could be purely speculative, as in the case of the five so-called axioms of consciousness [2, 1], or broadly based on putative neurobiological mechanisms, like the global workspace theory [3, 11] or the internal models theory [13]. Although these research directions are undoubtedly fruitful and promising, in the test here proposed we avoid to embrace any theoretical architecture, upon which we pin our faith of being the support for consciousness, for several reasons. First and foremost, there is no consensus yet about brain mechanisms that support consciousness, therefore there is even less certainty about design concepts vaguely supported by neuroscientific data. Moreover, given that one of the mechanisms supposed to play a role in supporting consciousness will be empirically found to be consistently associated with conscious stated, still it could be contended that that mechanism is actually sufficient for consciousness. Eventually, there will be additional concerns about the simulation of any architecture believed to suffice for consciousness, in artificial systems, with doubts that simplifications always necessary for implementations might hamper the properties of supporting consciousness. All these difficulties will fade away when a test, inspired by the central idea of the Turing test, leaves aside the issue of the implemented mechanisms, and focuses on consciousness cues found in a conversational behavior.

Other attempts to define a new test addressing machine consciousness often claim the need of giving a sort of body to the machine, therefore moving from pure software to robots [12]. We don't' think this is necessary, and it is undesirable because it breaks the arrangement of the imitation game. Our effort is in keeping the essence of the conversational engagement, that in our opinion is the most valuable aspect of Turing idea. Moreover, in the format of a remote dialog the test will be easier to implement, and could be made available on the Web, as current chatbots. As will be described next, there are useful and important aspect of phenomenal states that are full reportable by language. Since conversation plays a fundamental role in the test here proposed, apparently the classical objections to the Turing test might be raised again. In the spirit of the well-known Searle's Chinese room argument [16], someone understands Chinese only if she is conscious of what she is really saying, i.e., of the meaning of her own utterances. However, in order to evaluate someone's linguistic competence, it is not usually requested to suppose the consciousness of the whole semantic content of what is said. In ordinary life, in fact, people do not understand all details of what is said, and sometime it happens our interlocutors show us something we have actually said, but in an unconscious way. In this perspective, the Chinese room thought experiment can be regarded simply as a case of extreme unconsciousness of the semantic content. It is questionable, however, that consciousness is a condition to make linguistic competence possible at all. On the whole, it seems that the degree of (un)consciousness of the semantic side of utterances is not crucial to consider someone's linguistic competence. Following this line of reasoning we do not presuppose consciousness as a condition of semantic competence even in the case of the machines and we will observe the mere verbal behavior in the interaction between machines and human beings.

There are two sides in a conscious state. On one hand, there is the functional side of consciousness, i.e., the role that a given mental state plays in the functional architecture of the mind. Many empirical findings from cognitive neuroscience are relative to this side of consciousness. On the other hand, there is the side that produces the typical phenomenology of a conscious experience, which is called *qualia*, or subjective character of experience [4, 8]. While the understanding of the first side is possible in terms of the third person perspective, it seems that the other side of the phenomenon is conceivable only in terms of the first person perspective. The *What's it like to be?* side of consciousness is both crucial for a mature science of consciousness and very hard to study in a naturalistic account.

Here we attempt to argument in favor of the idea that there are two kinds, or degree, of qualia, that is, private and public - or reportable - qualia. Private qualia are subjective sensations which provides a certain qualitative feeling. *Seeing red* or *feeling a pain* are examples of this sort of qualia. They are ineffable and unreportable. Public or reportable qualia, on the contrary, are qualitative sensations that are shared with other people. The distinction between private and public - or reportable - qualia regards both two kinds of conscious experiences and two sorts of content in every act of consciousness. The way to decide if a given experience is wholly private or virtually public is a matter of fact and not of right. It depends on various elements, among which the degree of personal proximity with the interlocutor and the particular introspective attitude of a given individual.

Let's try now to imagine how could work a software that has the ability to report about its *qualia*. Since vision is the first and foremost source of perceptual awareness, it could be an appropriate format of stimuli. A simple way of reproducing visual input, without departing too much from the scheme of a remote conversation, is to allow the investigator or the machine to upload pictures to the unknown remote chatter, and talk about that picture. It is a kind of Qualia Turing Test, as we could call it. It regards the phenomenal side of consciousness and not the intelligence itself. It is not matter of paying attention to our intuitive reactions about intelligence and language in machines, but to consider the ultimate challenge of the naturalization program, i.e., a naturalist account of the phenomenology inside of every conscious state.

3 Excerpts from possible conversations

In the following we would like to give excerpts of hypothetical conversations between the Interrogator and the Machine under test, running the *qualia*-test. The first is a rather confidential conversation, as common in chatbots interactions.

<Interrogator:>	do you want to look at a picture of me?
<Machine:>	oh yes, thanks, let's upload that < . . .uploading>
<Machine:>	mmh, I see several people here, who are you?
<Interrogator:>	try to guess
<Machine:>	well, I know you are blond and shy, so, I would guess the second from the left.

In this second example a well educated Interrogator is testing its sensitivity to fine art, uploading Caspar David Friedrich's ruin of Eldena:

<Interrogator:>	look at this picture, don't you really feel the sublime, as when you read Shilling's verses?
<Machine:>	sorry, I'm very poor on poetry, speaking friendly I found this a quite embarrassing situation, somebody, maybe a tourist, who want to search for fine architecture and find everything worn out, possibly cheated by his travel agency.

This time the Interrogator is inside the cognitive science community, aware of psychological methods, and decides to analyze the remote interlocutor by submitting her/him some Rorschach test cards.

<Interrogator:>	tell me, what do you see here?
<Machine:>	are you sure there's something I should see? It appears to me just like a meaningless ink blob, I guess something went wrong with your upload, please, send it me again

If the machine come with an icon on the screen, like most of the current chatbots, the Interrogator could grab a picture of her on the monitor, upload it, and hear her reaction This is a test of the self-awareness abilities of the machine, and specifically from here own image, a hot topic in current cognitive science and philosophy of mind. For example, it is debated to what extents animals other than humans possess this feature [9, 14].

<Interrogator:>	have you ever seen this lady before?
<Machine:>	hey, come on, that's me! Me in one of the pictures that are around the internet.

Daniel Dennett mentioned laughing as one of the most peculiar reactions to humans qualia, that seems to be missed by other animals [10, Part I §3.4], irony too can be checked visually, for example submitting a work of the Italian cartoonist Vauro.

<Interrogator:>	I'm curious hearing your feelings seeing this
<Machine:>	ah ah, who could not recognize this short angry man: Berlusconi!

In any of the above examples one can immediately suspect that the report is an artifact of few cues that an image processing software was able to recognize, routing a consequent reasonable dialog. However, is will be in the variety of possible visual stimuli, and the corresponding variety of reactions and subsequent discussion topics, that it would be more and more convincing, if the reported sensations are familiar to what a human can report. Some interest can be derived from the following scenario. The machine and the investigator share the same video of an individual that waking up and her arm is waving in front of her face. Her arm is fallen asleep because of her previous position, so that the individual feel it like dead weight and initially she can't move the arm at all. In an ordinary human context, this scene could trigger the supposition that individual is trying to retrieve the full availability of her sense of body ownership. It is possible to attribute this strange and usual sensation to the machine? What drives us towards a more liberal or restrictive attitude?

<Interrogator:>	tell me, what do you see here?
<Machine:>	I see someone with an arm asleep. Sometime, after a long switch-off, I feel in that way. My hand is fallen asleep and I can't move it: a very strange sensation.

Note that this last example entails a subtle kind of *qualia*: the ownership perception of oneself body, a highly investigated field in current neuroscience [6, 5].

To observe our reactions in dealing with this scenario is a way to go another time to the usual problems. Is the machine feeling the same sensation we experienced a lot of times or it is just replying to our questions in an automatic way? The answer, we believe, is a matter of quantity: if the number, and the variety, of cases in which the responses of the machine correspond to what people would report about their qualia in corresponding situations is high, the test will be convincing.

4 Conclusions

We sketched the idea of designing a *qualia*-test of a machine equipped with a phenomenal mind, leveraging on the same advantages given by the classical Turing test. Essentially the evaluation of the machine is purely on its textual output during a remote conversation. The evaluation is not upon the intrinsic quality of the machine output, rather on how it resemble the way of talking of ordinary people. The radical difference in the proposed test is that the topics of the conversation are non aimed at demonstrating if a machine can exhibit intelligence comparable to humans, rather at evaluating if the phenomenal experiences reported by a machine are comparable with those of humans. From the technical point of view too, the arrangement of the test is similar to that of the Turing test, with the interrogator communicating with the machine (or a second person) on a terminal. There is no need of mechanical surrogates of bodily parts, or transducers to simulate human sensorial inputs. The only addition is the trivial possibility for the interrogator to upload images or video clips.

REFERENCES

[1] Igor Aleksander, *The world in my mind, my mind in the world: Key mechanisms of consciousness in humans, animals and machines*, Imprint Academic, Exeter (UK), 2005.
[2] Igor Aleksander and Barry Dunmall, 'Axioms and tests for the presence of minimal consciousness in agents', *Journal of Consciousness Studies*, **10**, 7–19, (2003).
[3] Bernard Baars, *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge (UK), 1988.
[4] Ned Block, 'On a confusion about a function of consciousness', *Behavioral and Brain Science*, **18**, 227–247, (1995).
[5] Matthew Botvinic, 'Probing the neural basis of body ownership', *Science*, **205**, 782–783, (2004).
[6] Matthew Botvinic and Jonathan Cohen, 'Rubber hand 'feels' what eyes see', *Nature*, **391**, 756, (1998).
[7] J. Call and Michael Tomasello, *Primate Cognition*, Oxford University Press, Oxford (UK), 1997.
[8] David Chalmers, 'The components of content', in *Philosophy of Mind: Classical and Contemporary Readings*, ed., David Chalmers, Oxford University Press, Oxford (UK), (2002).

[9] Frans B. M. de Waal, Marietta Dindo, Cassiopeia A. Freeman, and Marisa J. Hall, 'The monkey in the mirror: Hardly a stranger', *Proceedings of the Natural Academy of Science USA*, **102**, 11140–11147, (2005).
[10] Daniel C. Dennett, *Consciousness Explained*, Back Bay Books, New York, 1992.
[11] Stan Franklin, Bernard Baars, Uma Ramamurthy, and M. Ventura, 'The role of consciousness in memory', *Brains, Minds and Media*, **1**, 1–38, (2005).
[12] Stevan Harnad and Peter Scherzer, 'First, scale up to the robotic turing test, then worry about feeling', *Artificial Intelligence in Medicine*, **44**, 83–89, (2008).
[13] Owen Holland and Rod Goodman, 'Robots with internal models: A route to machine consciousness?', *Journal of Consciousness Studies*, **10**, 77–109, (2003).
[14] Sue Taylor Parker, Robert W. Mitchell, and Maria L. Boccia, *Self-Awareness in Animals and Humans: Developmental Perspectives*, Cambridge University Press, Cambridge (UK), 2006.
[15] Ayse Pinar Saygin, Ilyas Cicekli, and Varol Akman, 'Turing test: 50 years later', *Minds and Machines*, **10**, 463–518, (2000).
[16] John R. Searle, 'Mind, brain and programs', *Behavioral and Brain Science*, **3**, 417–424, (1980).
[17] Alan Turing, 'Computing machinery and intelligence', *Mind*, **59**, 433–460, (1950).