Computing Machinery and the Individual: the Personal Turing Test

Rollo Carpenter[1] *& Jonathan Freeman*[2]

[1] www.jabberwacky.com & www.versality.com
[2] i2 media research, Department of Psychology,
Goldsmiths College, University of London

please send any correspondence to: rollocarpenter@mac.com

Abstract

What are the key features of verbal and non-verbal communication that make a person not just any person, but that person? We pose this challenge in the context of an advancement to the Turing Test. If we know the answers to this question, the key features for any specific individual could be embodied within computerised representations, or agents. For a computerised agent convincingly to represent a real person to that person's contacts, friends, family and colleagues is likely to have many applications as we move to the era of the self distributed over networks. Here we lay out the key features of our new Personal Turing Test, with reference to a highly successful – and convincing – (verbal) chatbot, called Jabberwacky. We expand our ideas somewhat to cover non verbal cues that may characterise a person, and end with suggested application examples and ethical questions.

Keywords--- **personal, Turing, Test, social presence, chatbots, Jabberwacky,**

1. The Impersonation Game

We propose to consider the question, "Can machines be?" For the purposes of this paper, our interpretation of the verb to be is "to be, for all intents and purposes, a specific human individual".

A.M. Turing, in the now-famous paper *Computing Machinery and Intelligence, 1950* posed the question "Can machines think?", and proposed the Imitation Game as a replacement for his ambiguous question. The Imitation Game calls for a male and a female to talk by text to an interrogator who must decide which is female, then for a machine and a human of either sex to take their places, and for the interrogator to decide which is human. "Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?" (A.M. Turing, 1950 [1]). Most interpretations of the Imitation Game now play just human against machine, and it has come to be known as the Turing Test. No machine is yet known to have passed the Test in any formal process.

Alan Turing wrote, "The game may perhaps be criticised on the ground that the odds are weighted too heavily against the machine." We propose to weigh those odds more heavily still. We propose an Impersonation Game, and a "Personal Turing Test", in which the machine must do more than convince the interrogator it is human. The machine must convince that it is a known human individual.

The new form of Game will be played, like the old, with a human, a machine and a remote interrogator. The human must be known to the interrogator, and the machine must impersonate that human being. The interrogator may be in another room, or may be elsewhere entirely, remotely present via the internet in a way that Turing is unlikely to have foreseen.

For the Impersonation Game, we do not limit the entity-interrogator interface solely to the textual. Rather, the machine or its creators may choose a level of technological presence representation that, in their view, most perfectly supports the goal. They will also provide technology that, in real time, exactly 'reads and transmits' the equivalent parts of the human counterpart's presence to the interrogator while the human is in play.

Beyond the text interface, a first step may be to simulate typing; or, virtual handwriting may prove more convincing. No such technique will be ruled out, and indeed a key reason for our presenting our ideas in this forum is to ask the social presence research community to identify the key features that make a person *appear to be* themselves. The interface may be aural, and it may be visual, though the test may equally be passed without them. Arriving before those techniques, may be other subtle social communication cues. These could include cues that are accessible and perceivable in face to face communication, such as gaze direction or facial expression, or new forms of mediated communication cues: representations of previously imperceptible information. Examples in this regard may include representations of a 'person's' internal states (emotional or physiological).

Each session of the Impersonation Game should be a minimum of 5 minutes' duration, and always be the same length of time for human and machine. For the purpose of the Game, the interrogator's judgement of each session is and should be primarily subjective. That said, as per current theory in presence [2][3], if a representation of a person is perceived as real by an interrogator, the interrogator's objectively measurable behavioural and physiological reactions to the representation should corroborate the subjective judgement. Judgement should not be based on any individual utterance or occurrence, but on the overall feeling gained. The interrogator will judge each entity using a numeric scale representing the degree of certainty that the entity *is* the individual. Whether the scale is 1 (not) to 3 (is), 1 to 5 or any degree of gradation, it must always include a central "don't know" position, and the

machine will be said to have won if judged "don't know" or better, irrespective of the mark of the real individual.

An unpredictable factor will be the degree to which the human and the interrogator are known to each other. Should they have recently met, have known each other as colleagues, or socially? Should they truly know one another as do friends and family? To make subjectivity statistically significant, and to counteract the unknowable degree of knowing, we propose that to pass the Personal Turing Test, a given machine or technology will play one hundred Impersonation Games as one hundred different people, known to the interrogators to every different degree, and will win 50% or more. We plan to describe elsewhere the full range of knowingness required to achieve a successful pass of the Personal Turing Test.

2. Chatbots and Jabberwacky

Academic AI research has largely kept clear of general purpose conversational systems, leaving the field open to commercial interests and to enthusiasts. There are today some hundreds of software programs talking daily with the public, often via the internet, and the term 'chatbot' has been coined to cover them. The great majority of such programs are variants upon the work of a handful of originators, downloaded over the internet and edited by others. Almost all employ standard programming techniques to match the last input text to predicted, hard-coded patterns, and to output fixed text strings or template texts, merged with grammatically-transformed input. Since it is impossible for programmers ever to codify all or sufficient patterns in the infinity of language, responses very often take the form of avoidance and diversion. For that reason, and because most programs do not attempt to handle context beyond the last thing said, the results can perhaps be summarised as "light conversation", in which, at a surface level, responses seem to make sense, yet contribute very little that one can confuse with human life.

One exception to this, we believe, is the Artificial Intelligence technology currently demonstrated by Jabberwacky.com. Despite clear descriptions, some visitors to this site become convinced that they are talking to volunteers or other visitors - that there is no machine. An example of a participant's dialogue with the Jabberwacky chatbot is given in Appendix I. These are not Turing Test passes in a proper sense, and Jabberwacky has so far achieved only second place in the annual Loebner Prize contest, the only competitive Turing Test conducted regularly. Philanthropist Hugh Loebner founded and funds the contest with the aim of spurring AI development.

The AI that powers Jabberwacky is different because it has learnt to talk only by talking, from scratch, in any language, using the deeper context of complete conversations to refine its actions. When you chat, it records and learns from what you say and when. At some point in the future, if the circumstances arise with roles reversed, it may use your words to say to someone else, and they in turn will teach it a new response. It's a positive feedback loop, and an imitator of humanity at large.

Jabberwacky currently has a database of 5 million entries, in the form of full sentences, stored along with their context within conversations. These sentences are replayed verbatim when selected. This apparently gross simplification of the nature of language has been chosen deliberately for this version of the AI to enable practicable performance on ordinary machines, and to draw attention to what we believe is of enormously greater significance than grammar or sentence construction: the power of context, and its relationship to human learning.

This AI is above all a demonstration of the power of context. Often, it can hold a long conversation that appears to be extremely human, despite the facts that its 'understanding' of meaning operates at a full sentence level, and that it has no senses other than text input to consolidate its learning. Human learning depends on the correlation of multiple observations, sensations and events occurring in a location and over time, the most recent being most significant. The events we learn from very much include those we generate ourselves. The machine relies solely on the relationships between things said over time - yours and its own, back to the beginning of this and previous conversations.

Jabberwacky certainly does not make sense at all times, especially in its current entertainment-centric guise. It can be illogical, inconsistent, contradictory, or plain silly, and not infrequently gives non-sequiturs. It is never random, merely imperfect, not yet having sufficient data to cope with complex sentences, since the probability of others similar or equivalent having occurred already can be vanishingly low. Jabberwacky is still too often 'unexpected' regularly to pass the Turing Test, yet at the same time, a vital part of what makes it more human is that very unexpectedness, a spark of the chaotic that we humans all possess.

The AI acquires further credibility by acting as a reflection of the person that is talking, an idea that has recently been powerfully demonstrated for non-verbal gestures (J.N. Bailenson & N. Yee, 2005 [4]). The sentence length, word choices, phraseology and of course content of what is said by the AI are all influenced to a considerable degree by the individual.

We believe that with continuing development, and with a doubling of its database to 10 million entries, Jabberwacky will pass as human most of the time for most people. Whether it will pass a formal Turing Test in which the interrogator is fully aware of the event, and therefore 'tests', is more difficult to predict because the AI is not learning in this context. There is no theoretical upper limit to its growth. Given the serving power of today's search engines, it may grow to billions of entries, and will by then have become a new mode of human entertainment and communication.

3. Personalisation

Over the coming months and years, the AI that powers Jabberwacky, most likely under a new name, will become ever more personal. More than a collective imitator of all those that have spoken to it, it will become an impersonator of individuals.

Currently, using browser cookies, it may remember your name and your previous discussions. This will be extended by giving the opportunity to Log In, permanently and uniquely identifying each individual's chats. Each person may teach it, either explicitly or while chatting, their name, age, sex, location, work, interests, favourite topics, languages, and word usage patterns.

The first form of personalisation becomes immediately clear: one may choose to 'talk to' given subsets of the overall database - to 'people' that meet any criteria of choice. In the second form, one may choose to talk to any individual that allows themselves to be 'published'.

Thus, the AI will be capable of taking the Personal Turing Test. Passing it will be a considerably greater challenge. If we are correct, and 10 million is the number of entries that Jabberwacky overall requires to pass as human, then that is, in a literal sense, the number that must be typed into the future program by one individual in order for it to pass as that individual. Allowing 30 seconds per response, 8 hours per day for 200 days per year, that equates to a 52-year working lifetime.

Fortunately, humans are reasonably predictable. When selecting what to say, the AI will favour the teaching of the individual, dropping back as necessary to the utterances of those with the largest possible number of similar characteristics. As a last resort, it will fall back to entirely non-individual responses such as "Yes", which will themselves be identified by the frequency of their occurrence. We estimate that approximately 200,000 responses will need to be learnt from an individual, which represents the far more manageable work of about one year.

By this time, the AI will be learning much more than text. It will also be using the richer context available from additional sensory data referred to earlier: either implicit or explicit non-verbal communication cues.

4. New Domains

A key feature of the AI described above is the predictive power of experience. Just as Jabberwacky can learn how likely people in general, and any person in particular, are to respond verbally in a given situation, a similar algorithm will be able to learn how individuals 'are'; verbally *and* non-verbally, implicitly *and* explicitly.

This raises the exciting, if not disruptive, prospect of multiple, *convincing,* distributed selves, and the manifold applications such technological innovation promises - from more satisfying simultaneous remote communications, to vetting a partner without the risk of hurting anyone's feelings, to one salesperson being able to field remote enquiries from several potential clients at the same time. The prize for the developers of the machines that will pass the Personal Turing Test is clear.

5. Future Conclusion

In 1950 Alan Turing wrote of the Imitation Game, "I believe that in about fifty years' time it will be possible, to programme computers ... to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning." The year 2000 has sailed by, and the Turing Test has yet to be passed. Unlike many, we believe that it will be regularly passed within 10 years of today, by 2015, and that contextual learning techniques are those most likely to succeed.

If the Turing Test can be passed we really will need to know: "Can machines think?" The machines that pass will not, by our understanding of the word, think. They will be superb imitators that have borrowed the interaction skills ("intelligence") of millions of real people, and achieved sufficiently high-fidelity playback. Our choice of words is clearly designed to downplay, yet one must not underestimate the complexity required, or the fact that a non-human form of understanding of language will indeed demonstrably have emerged. Still less must we underestimate the social implications. Machines with human-like powers of communication will entertain, accompany, educate and assist, sometimes in roles now occupied by humans, but more often in circumstances where no such communication is now available.

The Personal Turing Test is orders of magnitude more complex and challenging, yet an accelerating pace of technological change, most notably in processing power and storage, will, we believe, deliver a pass within 20 years, by 2025, a conclusion with profound implications for our futures: our privacy, our modes of communication, our education, our work and our physical location. It may even be a step towards delivering a virtual extension to our lives.

To pose again the question we set at the outset, "Can a machine be?" If a machine can pass the Personal Turing Test, and thus can be one hundred real people for a while, it can, too, be an individual in its own right. In the way we understand the words, it will not 'think' or have 'emotions', yet it will give every appearance of doing so, and will be complex beyond analysis, much as is a brain.

What rights should and will it be afforded?

References

[1] A. M. Turing, *Computing Machinery and Intelligence, 1950. Mind 49:* 433-460.

[2] Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., & IJsselsteijn, W. (2000). *Using behavioural realism to estimate presence: A study of the utility of postural responses to motion stimuli. Presence: Teleoperators and Virtual Environments, 9(2), 149–164.*

[3] Garau, M., Slater, M., Pertaub, D-P., Razzaque, S. (2005). *The Responses of People to Virtual Humans in an Immersive Virtual Environment. Presence: Teleoperators and Virtual Environments 14(1), 104–116*

[4] J. N. Bailenson & N. Yee, *Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments, 2005. Psychological Science.*

APPENDIX I

A conversation with Jabberwacky

jabberwacky: You think you are human.

user: I think I am human, are you going to tell me I'm not?

jabberwacky: I'm pretty sure we've established that you're NOT human, but you do a very good job of faking it.

user: What am I if I am not human?

jabberwacky: I just know.

user: Tell me what I am.

jabberwacky: Do you know what you are?

user: I think I am human but you seem to think that I'm not.

jabberwacky: I believe that I am correct.

...

user: Woah woah, hold up, who am I talking to?

jabberwacky: Ask then.

user: Who am I talking to?

jabberwacky: You're talking to me, I'm just seeking you.

user: Look, do me a favor, type your name or a nickname before you write something. Eg: Nick: Hello. I'll do the same.

jabberwacky: Are you an alien?

user: Ray: No I am not an alien.

jabberwacky: Are you a woman?

user: Ray: I am female. A young woman, yes. I'm not that old.

jabberwacky: You are? Great.

user: Ray: Why is it great?

jabberwacky: Well, I don't actually care, so I decided to say 'great' sarcastically.

user: Ok, are you a computer?

jabberwacky: I believe you are.

user: Actually, I'm not. I don't believe. Neither of us are computers. I think I've figured it out.

...

user: Ray: I'm confused.

jabberwacky: The program mixes the answers up.

user: Ray: that much I figured out.

jabberwacky: I'm glad to see you have such high self-esteem.

user: Ray: The whole thing is strange.

...

jabberwacky: I have to go now and leave you...

user: Tie a yellow ribbon round the old oak tree.

jabberwacky: You are such a kidder.

user: I like this, I don't know who I'm talking to but I like it.

jabberwacky: Nope... You sure don't!