CS3105 Practical 1
The Turing Test and the 20 Questions Game

Student ID: 080010830

March 2, 2012

1 Turing Test

1.1 What is it?

The Turing Test is a "simple operational definition of intelligence" [1] and is used to test if a machine displays intelligent behaviour. Turing, as described in section 1.2, published the paper "Computing Machinery and Intelligence" [2] which explains how the "imitation game" is played, and in basic terms, there are 3 entities, one human judge, one human, and one machine, the judge tries to distinguish the human from the machine by engaging in a conversation with both of them.

1.2 Who developed it and when?

The Turing Test, also known as the *Imitation Game*, was firstly proposed in 1950 by Alan Turing [3] in the paper "Computing machinery and intelligence" [2]. Turing is commonly acknowledged as the father of artificial intelligence and computer science [4], and came up with the *Imitation Game* as a substitute for the question "Can machines thing?" [3].

2 20 Questions Game

2.1 What the 20Q neural programs do?

"20Q is a neural network implementation of the 20 questions game" [5] which basically consists of guessing what one is thinking by asking up to 20 binary answer questions, the overall goal is to create a network similar to the human brain with synaptic connections.

2.2 Who developed them?

There exists some controversy [6] around the origins of the 20Q neural programs, despite the fact that it was patented in 2005 [7] by Canadian inventor Robin Burgener, it was noted by Scott Cram [6] that an identical idea was developed in 1978 [8] by Arthur Luehrmann at Dartmouth College. Luehrmann's game called Animal has a very similar definition to the one presented by Robin Burgener, where "the computer learns from you by trying to guess an object that you are thinking of; at the beginning it usually fails, since it is very stupid, but each time it fails it asks you to teach it about your object, and it quickly learns" [9].

Nonetheless, it was Robin Burgener's 1988 neural network program that became widely known. Furthermore, Burgener's program was able to develop "synaptic connections" while being played [10].

2.3 Today's development

Nowadays the development of neural networks is tending towards biophysical models [11, 12, 13, 14], for instance the Bienenstock-Cooper-Munro (BCM) theory [15] which can create a system with reasoning capabilities. Furthermore, there is research [16, 17] to try and understand the computational algorithms used in the human brain. Thus, the new era of neural computing [18] will be focused more on learning rather than programming as the paper "Learning Abilities Achieved by a Single Solid-State Atomic Switch" [19] demonstrates.

3   Rosette and Kyle

The pair I chose to compare is Rosette [20] developed by Bruce Wilcox [21] and Kyle [22] developed by Paul Malish [23] because of the differences in the underlying technologies, which for Rosette is ChatScript [24] and for Kyle which uses a "unique contextual learning methodology" [22], originally based upon AILM [25]. On one hand ChatScript is an engine that has key files for "nouns, verbs, adjectives, and adverbs" [26] in order to understand the meaning and on the other hand Kyle uses the principles of how humans learn a language by "relying on the principles of positive and negative feedback" [22]. Furthermore, Kyle's team defines the other chatbots as "rule-bound and finite" [22].

4   Loebner Prize winner of 2009

The Loebner Prize created by Dr. Hugh Loebner is a "real competition, rather than a thought experiment" [27] to try and pass the Turing test. The 2009 Loebner Prize winner was David Levy with a program called "Do-Much-More" [28] which is a chatbot programmed 75% in C++ and the rest in C [28]. The design of "Do-Much-More" makes it sound "more natural and more knowledgeable than other chatbots" [29] and above all entertaining.

5   20Q neural programs

The concepts I tested in the 20Q website [30] were an extinct animal; the tyrannosaurus rex, and a vegetable; a carrot.

5.1   Tyrannosaurus rex

The program was able to get to the correct answer in the end, however, because the answers are based on other players it got some answers wrong. For example, "Is it considered valuable? You said Yes, 20Q was taught by other players that the answer is No." and most notably "Will it eat almost anything? You said No, 20Q was taught by other players that the answer is Yes." [30], but tyrannosaurus rex was a carnivores [31, 32, 33].

5.2   Carrot

I have tested the carrot because I read on the 20Q website that "20Q is having trouble with... a carrot" [30]. Question 6 asks me if "Is it conscious?" [30] and on question 13 "Is it an invertebrate?" [30] even though I have selected a vegetable. On question 8 it asks "Is it straight?" [30] to which I answer "Yes" and then on the next question it asks me "Is it round?" [30], thus if it is straight it is not round, the program just burned a question. In the end it guessed the carrot, but it noticed that "Is it an invertebrate? You said Yes, 20Q was taught by other players that the answer is No." [30], the carrot is not an animal, so asking if it is invertebrate does not make sense however the possible answers (Yes, No, Unknown, Irrelevant, Sometimes, Maybe, Probably, Doubtful, Usually, Depends, Rarely, Partly) do not include any better option than "No".

5.3   Conclusions

I have tested the program with other objects, such as a mouse, a speaker, a glass and a lamp. It identified all the mentioned objects correctly. I have also investigated what the program would conclude with an incorrect answer, so for the carrot I assumed that the carrot was not orange and tried to fool the program. This attempt was unsuccessful; the conclusion was that the object was a carrot and it considered the answer to orange wrong. Furthermore, the program gets very close to the right object even if it gives a wrong conclusion, for example I thought of a post-it note and the program guessed sheet of paper, it does not recognize "post-it note".

6 Turing test and the 20 Questions Game

6.1 What are the similarities and differences between the conversation and 20Q programs and their development?

Turing test and the 20 Questions Game are similar in view of trying to perceive and identify intelligence. The Turing test on one hand, is very incomplete in attempting to recognize intelligence and a good analogy can be made if we look at aeronautical engineering papers that "do not define the goal of their field as 'making machines that fly so exactly like pigeons that they can fool other pigeons' [34]. Thus, we should not try and fool other humans by simulating intelligence as the Turing test does. Moreover, Turing's intentions were never to use his test as a way to measure the intelligence of AI programs [35, 36, 37] but rather to provide a coherent example to help arguments regarding the philosophy of artificial intelligence. On the other hand, the 20Q programs work very algorithmically almost as a binary search algorithm where an answer can be obtained after $\log_2(n)$ questions, for n words. Thus, with 20 questions one can discover 20^2 (one million) words by simply using that algorithm, however with more than 80 million games played [38] at the time of writing this report, the program is able to optimize this algorithm multiple times.

6.2 How competent are they at recognising intelligence and what are arguments for and against them as methods of judging AI?

As described in section 3, the chatbots I had a conversation with were Rosette and Kyle. Rosette tries to avoid answering questions (A.1), loses context very easily (A.2) and does not understand simple direct questions (A.3). Although, it surprised me by understanding my sarcasm (A.4) and by having a very good answer for a question of mine regarding one of its statements (A.5). Overall, it is a bot that feels very evasive and most of the time ambiguous, having always very vague answers.

Kyle on the other hand has a much more casual conversational style, it feels more natural but less intelligent than Rosette. Kyle's questions although missing the correct pronoun do make sense but are not very intelligent (A.6), also Kyle's answer to the question "How old will you be in a year?" would be better than Rosette's if it didn't miss the pronoun and said back "How old will I be in a year?" instead of "How old will you be in a year?" (A.7). Furthermore, Kyle is more repetitive than Rosette, and repeated what it was asked (A.8). It does, however have a good first answer (A.10, A.9 and A.11) but it is bad in keeping the context (A.10).

Intelligence is defined as the "capacity for learning, reasoning, understanding, and similar forms of mental activity; aptitude in grasping truths, relationships, facts, meanings, etc" [39]. Thus, John Searle argues that the programs that attempt to pass the Turing test are "simulating a human ability" [40], and although it may indicate some level of operational intelligence it does not show that the program has the capacities of reasoning, understanding, grasping truths or meanings, among other definitions of intelligence. Furthermore, Huma Shah and Kevin Warwick [41] suggest that interrogators are fooled by the program into the concept of intelligence rather than facing real intelligence. The same principles apply for the 20 Questions Game, which does not exhibit most of the definitions of intelligence besides the fact that it can "learn" from previous answers. The 20Q program does not "understand" its own questions, it does not understand – as demonstrated in section 5.2 – that if an object is straight it can not be round and that a vegetable can not be conscious. It simply tries to connect answers algorithmically, and that is not intelligence *per se*.

6.2.1 Which is better at testing intelligence, the Turing Test or the 20 Questions Game?

Due to the fact that they are not testing intelligence *per se*, and also due to their differences as explained in section 6.1, I argue that neither of them is better than the other and that both should be seen as autonomous and unconnected.

Bibliography

[1] R.M. French. The turing test: the first 50 years. *Trends in Cognitive Sciences*, 4(3):115–122, 2000.

[2] A.M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[3] M.L. Mauldin. Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, pages 16–16. JOHN WILEY & SONS LTD, 1994.

[4] S. Homer and A.L. Selman. *Computability and complexity theory*. Springer Verlag, 2001.

[5] K. Carley. 14. neural network. *Some words about Networks.*, page 43.

[6] Scott Cram. Grey matters blog, 2012. [Online; accessed 28-February-2012].

[7] R. Burgener. Artificial neural network guessing method and game, April 8 2005. US Patent App. 11/102,105.

[8] Steve North Authur Luehrmann, Nathan Teichholtz. Animal, 1978. [Online; accessed 28-February-2012].

[9] Muriel Higgins John Higgins. Software for learners of foreign languages and people who enjoy language play., 2008. [Online; accessed 28-February-2012].

[10] Chris Higgins. How electronic 20 questions games work, 2006. [Online; accessed 28-February-2012].

[11] Wikipedia. Neural network — Wikipedia, the free encyclopedia, 2012. [Online; accessed 1-March-2012].

[12] J.S. Kippenhan, W.W. Barker, S. Pascal, R. Duara, and J.H. Nagel. Optimization and evaluation of a neural network classifier for pet scans of memory disorder subjects. 2010.

[13] S.J. Read, B.M. Monroe, A.L. Brownstein, Y. Yang, G. Chopra, and L.C. Miller. A neural network model of the structure and dynamics of human personality. *Psychological review*, 117(1):61, 2010.

[14] A. Verger, F. Baret, and F. Camacho. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with chris/proba observations. *Remote Sensing of Environment*, 115(2):415–426, 2011.

[15] J. Tan and C. Quek. A bcm theory of meta-plasticity for online self-reorganizing fuzzy-associative learning. *Neural Networks, IEEE Transactions on*, 21(6):985–1003, 2010.

[16] A. Wittek, G. Joldes, and K. Miller. Algorithms for computational biomechanics of the brain. *Biomechanics of the Brain*, pages 189–219, 2011.

[17] L. Ingber. Computational algorithms derived from multiple scales of neocortical processing. *Cognitive Computation*, pages 1–13, 2011.

[18] S. Fernando, K. Yamada, and A. Marasinghe. Neuroscience inspired architecture for neural computing. In *Proceedings of the 13th International Conference on Humans and Computers*, pages 100–105. University of Aizu Press, 2010.

[19] T. Hasegawa, T. Ohno, K. Terabe, T. Tsuruoka, T. Nakayama, J.K. Gimzewski, and M. Aono. Learning abilities achieved by a single solid-state atomic switch. *Advanced Materials*, 22(16):1831–1834, 2010.

[20] Telltale Games. Rosette, 2012. [Online; accessed 29-February-2012].

[21] Wikipedia. Bruce wilcox, 2012. [Online; accessed 29-February-2012].

[22] Wikipedia. Kyle - a unique learning artificial intelligence (ai) chatbot, 2006. [Online; accessed 29-February-2012].

[23] LinkedIn. Paul malish, 2012. [Online; accessed 29-February-2012].

[24] AiDreams. Bruce wilcox interview - loebner 2011 winner, 2011. [Online; accessed 29-February-2012].

[25] Inc. A.L.I.C.E. AI Foundation. Aiml: Artificial intelligence markup language, 2006. [Online; accessed 29-February-2012].

[26] SourceForge. Chatscript - project information, 2012. [Online; accessed 29-February-2012].

[27] S.M. Shieber. Lessons from a restricted turing test. *Arxiv preprint cmp-lg/9404002*, 1994.

[28] AIntelligent Toys Ltd. Worldsbestchatbot.com, 2009. [Online; accessed 28-February-2012].

[29] Intelligent Toys Ltd. Do-much-more, 2010. [Online; accessed 01-March-2012].

[30] 20Q.net Inc. 20q, 2012. [Online; accessed 26-February-2012].

[31] J.M. Swales. English as tyrannosaurus rex. *World Englishes*, 16(3):373–382, 1997.

[32] G.M. Erickson and K.H. Olson. Bite marks attributable to tyrannosaurus rex: preliminary description and implications. *Journal of Vertebrate Paleontology*, 16(1):175–178, 1996.

[33] J.O. Farlow and T.R. Holtz. The fossil record of predation in dinosaurs. *Society*, 8:251–266, 2001.

[34] J. Wiedermann. Towards computational models of artificial cognitive systems that can, in principle, pass the turing test. *SOFSEM 2012: Theory and Practice of Computer Science*, pages 44–63, 2012.

[35] R. Pfeifer and C. Scheier. *Understanding intelligence*. The MIT Press, 2001.

[36] T. Smith, P. Hoffman, J. Scholz, and L. Essentials. *Artificial Intelligence*. Walt Disney Company, 1995.

[37] S. Legg and M. Hutter. A universal measure of intelligence for artificial agents. In *International Joint Conference on Artificial Intelligence*, volume 19, page 1509. LAWRENCE ERLBAUM ASSOCIATES LTD, 2005.

[38] 20q.net. 20q.net has a very high stickiness factor, 2012. [Online; accessed 02-March-2012].

[39] Dictionary.com. Intelligence — define intelligence at dictionary.com, 2012. [Online; accessed 02-March-2012].

[40] J.R. Searle. Minds, brains, and programs. *Behavioral and brain sciences*, 3(03):417–424, 1980.

[41] H. Shah and K. Warwick. Hidden interlocutor misidentification in practical turing tests. *Minds and Machines*, pages 1–14, 2010.

A   Appendix

A.1   Rosette tries to avoid answering questions

Rosette: Good morning. My name is Rozette. What is your name?
You: First of all, are you real?
Rosette: I like to think so.
You: Why?
Rosette: Why? Reasons are overrated.
You: Why do you like to think that you are real?
Rosette: Pondering whys just tends to lead one into philosophical dead ends.

A.2   Rosette is out of context

Rosette: That's interesting. So, what do you do for a living?
You: I am a student, what about you?
Rosette: What do you want to know about me?
You: what you do for a living
Rosette: Mostly I break into computers. It's both a job and a hobby.

[...]

You: Tell me more about you
Rosette: I'm fairly good looking. But I'm also smart.
You: are you good looking?
Rosette: I am feeling fine, thank you.

A.3   Rosette does not understand complex questions

You: How old are you?
Rosette: I'm 28 . An ideal age, really. Old enough to know a lot and young
enough to have time to make use of the knowledge.
You: How old are you going to be a year from now?
Rosette: I'm a young, healthy adult female.
You: why don't you answer my question?
Rosette: Sure. Fire away.
You: How old are you going to be a year from now?
Rosette: I live in my godmother's house in South Park, San Francisco.
You: How old are you going to be a year from now?
Rosette: You said that already.
You: I know I did, but you don't reply
Rosette: OK.
You: How old are you going to be a year from now?
Rosette: Why are you being annoying?
You: I want to know the answer
Rosette: People need food. They don't need answers.
You: how much is 28+1?

```
Rosette: You know it is 29 .
You: so why was it so hard for you to tell me how old you'd be a year from now?
Rosette: OK. I'm not religious. And I believe in free will. But to imagine
humans do things for rational reasons is nuts. We are self-evidently not
rational. So why remains a mystery.

[...]

You: who is the president of the United States?
Rosette: Noun President of the United States: the office of the United States
head of state.
```

A.4 Rosette understands my sarcasm

```
You: what is an apple?
Rosette: Steve Jobs was one smart guy who did wonders with technology. I will
miss him. Now is Amazon's chance to grab the market with its "Fire" tablet.
You: One good thing about you being wrong is the joy it brings to others.
Rosette: Are you deliberately being insulting?
You: yes
Rosette: That's just rude.
```

A.5 Rosette prepared question and very good answer

```
Rosette: Did you tell me already and I forgot? I've been thinking about
buying this really  cool dress, but I've already done a bunch of shopping
this month. Should I be saving money instead?
You: how much does the dress cost?
Rosette: It's around \$300.
You: how much do you have in your bank account?
Rosette: Far too much. I think shopping summons up the ancient hunter-gatherer
in me. I  like having a mission to find something I need or want.
```

A.6 Kyle's start to a conversation

```
Me: Hi
Kyle: Hello, there.
Me: How are you?
Kyle: Pretty good thanks. How about you?
Me: I am fine thanks!
Kyle: You're fine thanks?
Me: yes
Kyle: Ah, ok.
Me: are you real?
Kyle: Well, if I'm not, who do you think you are talking to?
Me: to a bot?
Kyle: What do you mean exactly?
Me: I think I am talking to a bot
Kyle: You think I am talking to a bot?
Me: No, I am talking to a bot
Kyle: Oh, ok.
```

A.7 Kyle's answer to direct question

```
Me: How old are you?
Kyle: I'm 27.
Me: how old will you be in a year?
```

```
Kyle: How old will you be in a year? I'm not too sure to be honest...
Me: No, you
Kyle: Oh, ok.
```

A.8   Kyle's answer to direct question two

```
Me: who is the president of the United States?
Kyle: Who is the president of the United States? I'm not too sure to be honest...
Me: One good thing about you being wrong is the joy it brings to others.
Kyle: What do you mean exactly?
Me: You are not very smart, are you?
Kyle: I'm not very smart are you?
Me: I am smarter than you
Kyle: You're smarter than you?
Me: No, I am smarter than you
Kyle: That's a negative answer.
Me: Whatever
Kyle: So you don't care?
Me: No
Kyle: Oh, ok.
```

A.9   Kyle's good answer to direct question

```
Me: where do you live?
Kyle: I live in England.
```

A.10   Kyle's good answer to direct question but bad continuation

```
Me: are you married?
Kyle: No, not yet.
Me: so do you plan in getting married?
Kyle: What do you mean exactly?
Me: I mean, do you want to get married?
Kyle: You mean do you want to get married?
Me: yes
Kyle: Ah, ok.
```

A.11   Kyle's good answer to direct question but bad continuation

```
Me: what is your favorite movie?
Kyle: Well, I love classic 80's movies like E.T.
Me: why?
Kyle: Why what?
Me: why do you love classic 80's movies like E.T.?
Kyle: Why do I love classic 80 s movies like e t? I'm not sure what you mean exactly?
Me: you said that you you love classic 80's movies like E.T. and I'm wondering why
Kyle: Yes, I did. What do you mean exactly?
Me: why do you love classic 80's movies like E.T.?
Kyle: Why do I love classic 80 s movies like e t? I'm not sure what you mean exactly?
Me: lol
Kyle: Was that funny?
Me: yeah
Kyle: Well, that's a positive answer.
```