

Approaching Artificial Intelligence for Games - the Turing Test revisited

Jenny Eriksson Lundström and Stefan Karlsson

*Department of Information Science, Computer Science Division
Uppsala University, Sweden
jenny.eriksson@dis.uu.se, stefan.karlsson@dis.uu.se*

Abstract: Today's powerful computers have increasingly more resources available, which can be used for incorporating more sophisticated AI into home applications like computer games. The perhaps obvious way of using AI to enhance the experience of a game is to make the player perceive the computer-controlled entities as intelligent. The traditional idea of how to determine whether a machine can pass as intelligent is the Turing Test. In this paper we show that it is possible and useful to conduct a test adhering to the intention of the original Turing test. We present an empirical study exploring human

discrimination of artificial intelligence from the behaviour of a computer-controlled entity used in its specific context and how the behaviour responds to the user's expectations. In our empirical study the context is a real-time strategy computer game and the purpose of the AI is merely to pass as an acceptable opponent. We discuss the results of the empirical study and its implications for AI in computer applications.

Keywords: The Turing Test, Imitation Game, AI, Game AI

1 Introduction

Artificial intelligence in home applications has been reoccurring as a utopian idea for a very long time. So far, though, we have not seen much in the way of commonly used applications drawing on artificial intelligence. One area that has consistently tried to use AI ideas is computer games. From the simple pattern-based movements of *Space Invaders* (Taito, 1978) or *Pacman* (Namco, 1980), through the emergence of game-AI in finite state machine controlled behaviour, e.g. *Doom* (Id Software, 1993), to *Black and White's* (Lionhead Studios, 2001) in-game creatures that use reinforcement and observational learning to bring about highly emergent behaviours (Schwab, 2004). However, throughout the history of computer games there has always been a shortage of CPU power resulting in AI programming limitations. As a consequence, the challenge presented by computer-controlled entities never compared to those controlled by humans. This is changing. Today's powerful computers have increasingly more resources available, which can be used for incorporating more sophisticated AI into the computer games. But what is really needed?

1.1 Game AI and AI for games

According to Russel and Norvig (Russel and Norvig, 1995) "AI is the creation of computer programs that emulate acting and thinking like a human, as well as acting and thinking rationally." When it comes to defining AI used in games, even though broader definitions exist, game AI is primarily concerned with how

the system acts and leaves little interest in how the result was obtained. "[...] game AI is the code in a game that makes the computer-controlled opponents (or cooperative elements) *appear* to make smart decisions when the game has multiple choices for a given situation, resulting in behaviours that are relevant, effective, and useful." (our italics) (Schwab, 2004). Therefore, it is clear that adding more sophisticated AI to a game is only of interest if the result enhances the game experience.

The perhaps obvious way for AI to enhance the experience of a game is to make the player perceive the computer-controlled entities as intelligent. This brings us to the question of what we humans consider intelligent. Through history, the qualification of intelligence has changed. As an example, it was long thought that our human ability to use tools set us apart from animals, thus providing the qualification for intelligence. When this was proved wrong, the qualification was altered. Other discriminators that have been used include the use of language, and sense of time. It seems clear that we humans want to be the only 'truly' intelligent species. However, in some cases we are more than willing to make exceptions. For example, we are much more likely to attribute intelligence to a beloved pet. Does the same hold true for the computer-controlled entities with which we now interact?

1.2 The Turing Test

The traditional idea of how to determine whether a machine can pass as intelligent is the Turing test. The Turing test was first described by Alan Turing in his article *Computing machinery and intelligence*, *Mind*, 1950.

1950-1966	A source of inspiration for all in AI
1966-1973	A distraction from more promising AI
1973-1990	Source of distraction to philosophers
1990-	Consigned to history?

Figure 1. The common view on the significance of the Turing Test according to Whitby (1996)

Since then, there has been serious criticism of the test, which has resulted in several suggestions for modifications, as well as changing the ascribed significance of the test itself (see Figure 1). The questions regarding the validity of the test posed by Marvin Minsky (Minsky, 1982) sum it all up. Would it even be possible for any machine to pass such a test? More importantly: Do passing the Turing Test constitute a sufficient condition for intelligence or not? What does this mean for our questions on whether humans are willing to accept computer-controlled entities as intelligent? One interesting observation concerns ELIZA (Weizenbaum, 1966). It has been shown that people have accepted chat-bots like ELIZA as human for a period of time. But more importantly, people kept talking to ELIZA even after it was revealed that she wasn't human but a program. It seems as if people are willing to set aside their knowledge that the counterpart is artificial and accept certain programs as, if not intelligent, at least as personalities. So, what makes people willing to set aside their knowledge that the entity is artificial and not human? Is it ultimately a matter of knowing but still accepting?

1.3 Testing the 'untestable'

In a game setting, players almost always know whether an entity is human or computer-controlled. However, players first start to complain about the entity when it does not live up to their expectations of its behaviour, or when they believe that it breaks the rules of the game. The expectations on the entity differ according to the significance of the role in the game the entity is to take. Still, the baseline of these expectations is usually what the players themselves would do in the same situation, or to generalise, what a human would do. Finding a way to evaluate how well a computer-controlled entity fulfils its role would be useful for design and utilization of AI in games and other similar applications. With a method for AI

evaluation, the adequacy of settings in simulation-based training as well as the creation of stimulating computer-controlled opponents would be enhanced. As earlier stated, in most games the only observable qualities of an entity, human or artificial, is the opponent's behaviour. Consequently, only the behaviour of the AI is available as a base for determining whether or not a computer-controlled opponent is accepted as a good opponent. To evaluate a computer-controlled entity would mean to compare its actions against those of a human in the same role.

2 Turing's test ≠ The Turing Test?!

In his 1950 paper, Alan Turing claims that the question "Can machines think?" is not a feasible question to pursue in order to establish whether machines possess intelligence. Instead, Turing offers a more feasible benchmark; illustrated by the description of a game he calls "The Imitation Game".

2.1 The Imitation Game

Turing (Turing, 1950) describes the Imitation Game as follows:

It is played by three people, a man(A), a woman(B), and an interrogator(C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. ... We now ask the question, 'What will happen when a machine takes the part of A in this game? 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?' These questions replace our original, 'Can machines think?'

2.1.1. Why the Imitation Game?

By choosing a game setting, Turing restrained the interaction possibilities and the relevant context to prevent prejudices and preconceptions, and thus tried to provide equal conditions for the man and the machine. Discussing the test setting, it is important to recognize that even though it is the method of communication in the test, the test itself is not a test of natural language skill. Today, even though natural language processing is still a disability for machines compared to humans, interaction with other entities, machine or human, can be made in virtual environments and thus, the necessity of using natural language as a base for the observations no longer exists. In accordance to Turing's idea, by using a virtual environment, the assessment of the machine's performance against a human benchmark could now be conducted with a level playing field.

2.1.2. How should the Imitation Game be interpreted?

One interpretation of the Turing Test is as a test of whether machines can fool humans. Taking into account the absurd settings presented in various 'Candid Camera inspired' shows, as well as the experiences with ELIZA and other chat bots, it seems clear that given the right circumstances humans are very susceptible to be fooled. Consequently, even though of philosophical interest, to test whether a computer could fool an observer into thinking that the computer is a human, has limited practical application. What is needed is not confirmation of whether there are intelligent or deceiving machines. Marvin Minsky (Minsky, 1982) expressed that: "The very essence of the Turing test is our judgment of how well machines *act* like humans." (our italics). We claim that it is more interesting to see whether the machine behaviour and performance measure up in a particular situation. Thus, our interpretation of the Imitation Game is whether a machine could replace a man as A, and *judged by* a human observer C, *be equally successful as the man in the role of A*. This question of how and how well the entity fulfils our expectations on performance in a given situation could prove to be highly useful for evaluating and assessing computer performance in a role normally played by a human. And given the human inclination

of using ourselves as baseline, is crucial for the development, creation and utilisation of computer controlled entities.

3 The empirical study - set up

An empirical study could provide useful information on whether a variant of the Turing Test is possible and useful for AI in games as a test of quality and relevance of use. AI in games is about "acting human" rather than making rational decisions. Also, the primary (and perhaps only) way to observe whether AI in games fulfil their roles is to observe their actions. In our empirical study the context is a real-time strategy computer game, *Dawn of War* (Relic, 2004), where computer-controlled players take the same role as human players. The purpose of the AI is merely to pass as an acceptable opponent. For the study a circular map divided into six pie-like sections was chosen. The three participants (the AI, the human control player and the test player) each got a fixed start section, and the game was set to a simple three-way battle.

Seven experienced computer game players participated in the study. First the players were to familiarise themselves with the game by playing the tutorial. Then they played one game against two opposing players, of which one was human-controlled and the other was computer-controlled. In order to circumvent the automatic labelling of the computer-controlled player as "computer 1" from jeopardising the study, the human-controlled player was labelled "computer 2". No participant commented on or seemed to notice this. It should be noted that the human opponent was a very skilled player while the computer-controlled opponent was of standard difficulty for the game and thus turned out to be mediocre in comparison. After the session, the participants completed a questionnaire on their both opponents, including stating and motivating which opponent was human/AI and what the participants considered to be good/bad with each opponent.

4 Implications of the study

Given the questionnaires, all participants had an opinion on how AI and human opponents behave and believed that they possessed the capability of making this discrimination. Five out of seven participants correctly identified which of the opposing players was human and which was computer-controlled. For those two that made an incorrect identification, they gave a concrete action/incident as the motivation to why they believed the computer controlled entity to be controlled by a human. When asked the participants what they thought signified a good opponent, an extra dimension to good opponent were brought up, namely fun to play against. The participants explicitly stated that this included that the opponent should be about evenly matched with the player, shouldn't exploit rules or otherwise cheat. Interestingly, cheating has historically been a common way to increase the challenge posed by computer-controlled opponents (Rouse, 2005). For the most part, all participants had elaborate motivations that seemed grounded in the behaviour of the opponents and the participants' earlier experiences with human and computer-controlled players. The participants demonstrated a complex chain of reasoning on consequences, behaviour and expected performance. They seemed to base their judgement on previous experience with how human and computer opponents behave in a given role, and their views of how a human respective a computer controlled entity would perform and take on that role. Thus, even though two participants were incorrect, it could be said that the participants seem to be able to discriminate AI from the human in the role as opponent in this strategy based game and uses concrete actions and that they use behaviour as the basis of discrimination.

5 Conclusion

The study shows that a variant of the Turing Test performed on actions is both possible and useful. The participants were able to form an opinion on each opponent and based this evaluation of how well the agent performed in its role on observations of concrete actions. Such opinions are clearly valuable for design and incorporation of AI in games since the participants' evaluation of how well the AI performed in

its role incorporate the comparison to the behaviour of a human opponent and express the participants' expectations and level of acceptance.

6 Thoughts

As argued in this paper, it is possible to conduct a test adhering to the intention of the original Turing test. The important thing for resolving this question is that humans are able to reason based on the observed behaviour. For this purpose, whether the participants were right or wrong is a non-issue. However, one cannot fully ignore the two participants making the incorrect identification and its implications. So, what caused these two to make a wrong identification? From the answers in the questionnaires, we can see that the main reason was that the participant observed a concrete action which he/she saw as a crucial mistake. Behaviour the participant considered that a human player never would have demonstrated. In other words, the action was interpreted as a result of 'non-human' reasoning. Along the same line, the actual computer-controlled player (regardless if it was identified as such) was described as unskilled and lacking knowledge about the situation and the worth of its resources.

Humans seem to have a firm belief in the own capabilities of discrimination of artificial intelligence from humans. Even though now challenged, the fact that no computer program so far has been considered to fool an 'informed' human enough to receive the Loebner Prize gold medal implicates that the belief in the capability of humans to make this identification, still is substantially unshaken. Given this, one important question comes to mind: If humans always are able to discriminate AI based on actions/behaviour, the consequences of substituting humans for AI in the role of humans must be questioned for training and simulation situations. The reason for why this kind of training cannot be adequate is because this set-up does not train humans to face and handle human behaviour. We only learn how to handle a machine trying to take on the human role. That is, at best we are training according to the rule-based implementation of someone's informed opinion of how a human would behave. Therefore, we might only learn to relate to optimised logical and 'rational' rule-bound reasoning and not to reasoning affected by emotions, loyalties and other motivations beyond rule-bound rationality. In all, only learn interaction patterns that are not considering the very essence of what separates us humans from machines.

References

- Id Software (1993) *Doom*.
- Lionhead Studios (2001) *Black and White*.
- Minsky, M. (1982) Why People Think Computers Can't. *AI Magazine* 3(4), pp. 3-15.
- Namco (1980) *Pac-Man*.
- Relic (2004) *Dawn of War*.
- Rouse, R. (2005) *Game Design: theory and practice*, 2nd ed. WordWare Publ., Plano, Texas.
- Russel, S. and Norvig, P. (1995) *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs, New Jersey.
- Schwab, B. (2004) *AI Game Engine Programming*, Delmar Thomson Learning.
- Taito (1978) *Space Invaders*.
- Turing, A. M. Computing machinery and intelligence, *Mind*, vol LIX, no 236, pp. 433-460.
- Weizenbaum, J. (1966) ELIZA--A Computer Program For the Study of Natural Language Communication Between Man and Machine, *Communications of the ACM*, Volume 9, Number 1 pp. 36-45.
- Whitby, B., (1996) "The Turing Test: AI's Biggest Blind Alley?" in P. Millican and A. Clark (eds.) *Machines and Thought: The Legacy of Alan Turing*. Volume 1, Clarendon, Oxford.