On the Perceived Intelligence of Personified Agents

Matthew Bedder
University of York
York, United Kingdom
mb708@york.ac.uk

ABSTRACT
The goal of much of the current research into Artificial Intelligence is to generate AI techniques that allow computers to be able to play games at or above the level of humans. A secondary and potentially underexplored goal in generating AI systems for games tries to create techniques to allow computer players to appear intelligent, but rather use unfair advantages in order to remain competitive. For this secondary approach it may be useful to consider how we can increase the perceived intelligence of artificial systems.

For this assessment I describe an experiment that tests whether players' perception of the intelligence of artificial agents is affected by the giving them human-like appearance. In a small study ($N = 13$) I am unable to prove that this is the case, but further propose more experiments that can be performed to better understand whether visual representation of AI players impacts perceived intelligence.

Author Keywords
Digital games, Artificial Intelligence (AI), player perception, personification

INTRODUCTION

Motivation

Much of the research performed in the area of Artificial Intelligence for games focusses on increasing the performance of AI agents. Some games, however, remain too complex to be solved efficiently for even state-of-the-art AI techniques, and so the problem of how to deal with these games remains.

It may be argued that, rather dedicating so much research towards generating stronger AI, more effort should be put into looking for ways to generate interesting AIs, or AIs that are fun to play against. Johnson argues that using simple AIs that "cheat", being afforded better powers or abilities than human players, could allow difficult games such as Civilization to have AIs that are both enjoyable to play against and provide a sufficient challenge [1].

It has been noted, however, that players of games may not enjoy games where it is obvious that the AI is cheating [2]. For cases where AIs have to cheat in order to provide a sufficient challenge for humans it could therefore be useful to be able to make cheating AIs to be able to appear more intelligent than they actually are.

Existing Research

Work by Denisova and Cairns [3] suggests that the priming of players' expectations about the complexity of AI systems can in fact impact the perceived intelligence of those systems. In their experiment participants were made to play a game twice, once where they were told that they were playing the game's standard AI and once where they were playing "an adaptive AI". Despite no actual changes being made to the game it was found that players became more immersed in the game with "adaptive AI", regardless of whether the player had any prior experience in games with adaptive AIs. Quotes from [3] also seem to suggest that participants considered the supposedly different AI to be more intelligent, and provide a more tailored experience.

Whereas this approach requires players to be told that the AI is stronger than it actually is we can consider approaches where agents are made to seem more human-like in order to try to increase their perceived intelligence.

Koda and Maes tested whether the visual representation of artificial agents in a poker game impacted player perception of the agents' intelligence and likability [4]. In one experiment participants were made to play against two artificial players, with the first having a "Caricature Human" face and one having no visual representation, with no significant difference being found in the perceived intelligence of the two agents. It may, however, be argued that this experiment contained two major flaws that may have contributed to this result: firstly, the "Caricature Human" face was animated in such a way that it could give away whether the agent was bluffing, meaning that the human player would be able to get more information from the "Caricature Face" agent than the agent with no representation, potentially reducing the perceived intelligence of the "Caricature Face" agent somewhat; secondly, participants were made to play against both agents at the same time, allowing direct comparisons to be made between the agents, and making it more likely that the participant would realize that both agents played similarly.

A further experiment performed in [4] had participants playing against three agents, one represented by a crude smiley face, one with a "Caricature Human" face, and one with a more realistic face. In this experiment Koda and Maes did notice a slight increase in the perceived intelligence of the more realistically drawn agents, although not in a significant way. This experiment also suffered from allowing participants to directly compare the actions of the tested agents, but still highlights that there may be some potential for a link between how human-like an agent is portrayed and how its intelligence is perceived by players.

McMahan argues that the realism of a game can be a factor that affects the immersive qualities that a game may possess [5] with realism being broken down into "perceptual realism", or how realistically are objects depicted, and "social realism", or how realistically interactions and relationships depicted. If we consider that the absence of both social and perceptual realism may cause reduced immersion in players we may also presume that the lack of these realisms may reduce the ability of players to consider artificial agents as intelligent.

EXPERIMENT

As it is apparent that there may be some link between how realistically an agent is portrayed and the perceived intelligence of said agent it may be useful to investigate this link further, with the aim of knowing whether efforts to personify agents can be used to increase their perceived intelligence.

The experiment designed for this assessment therefore aims to look at whether a player might consider adversarial agents in a simple video game more or less intelligent based on how human-like they are portrayed.

Hypothesis

I propose that if two variants of the same game are presented to two sets of participants, with the variants differing only in how human-like the agents in the game are portrayed, that the participants playing the version of the game with the personified agents would consider them more intelligent than the participants playing the variant with non-personified agents.

Experiment Design

A number of participants were asked to play one of two variations of a game, one with non-personified units and one with personified units, and were then asked to complete a questionnaire about their demographic and experience with the game, with a particular focus on their opinion on the intelligence of the enemy units. Gameplay data on the amount of levels completed by the participant and on the number of times the participant failed was also recorded using the logs from the game.

The questionnaire used consisted of fourteen questions, of which the first seven where on the demographics of the participants and was taken from the IEQ questionnaire used in the RIGI module[1]. The eighth and ninth questions in the questionnaire asked about how much the user enjoyed the game and how difficult they found the game, while the remaining five questions asked about aspects of the AI seen in the game.

As participants would likely notice that there was no difference in the enemy intelligence between the two games, or would ascribe the intelligence of the enemies on how hard the levels played on each variant, a between-participants experiment design was used. In order to reduce the boas introduced by either the experimenter or by external factors (such as the time of day) identical instructions were given to participants before being shown the game, and the game's graphics would be pseudo-randomly selected just before they started playing.

The questions used to collect this data can be found in the Appendix at the bottom of this submission

Experiment Procedure

Participants were first given a consent form explaining the process they would undertake. They would then play of one of two game variants on a laptop for five minutes, or until the participant completed all of the available levels, or asked to stop playing. The participant was then asked to complete a short questionnaire. Once the experiment was over, each participant was debriefed and given the opportunity to ask any questions they had about the game.

A second form was later filled in containing details from the games logs on metrics of the gameplay session.

Resources Used

For this experiment a simple turn-based game (hereon referred to as HexGame) was developed in order to test graphics styles where the characters are and are not represented as humans[2]. This game, partially based off the mobile game Hoplite [6], involves moving a player-controlled unit around pre-designed levels in order to defeat enemy units.

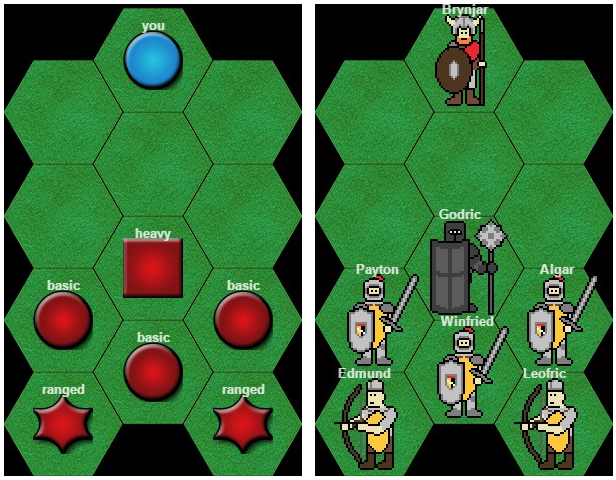

Figure 1: Level 14 of HexGame for both graphical styles used in the experiment (non-personified on the left, personified on the right)

[1] A small edit to the question about the age of the participant was made to remove values duplicated across different categories.

[2] The source code for HexGame and both playable versions of HexGame used in this experiment can be viewed at http://www.bedder.co.uk/iggi/rigi/1/

This game was designed with simplicity in mind, and as such the only action that the player can perform is selection of which tile to move to, with attacks between the player's character and enemy units being performed based on the relative locations of each other. In order to reduce the number of instructions that would have to be given to each participant the first levels of HexGame are used to teach the player about the game's mechanics.

In one version of the game you player-controlled character was represented as a Viking with the name "Brynjar" displayed above their head. Enemy units are displayed as knights, heavy knights, and archers, and each have an archaic name displayed above their heads (such as "Payton", "Edmund", and "Godric"). In the second variation of the game the playable character is represented by a blue circle, and the enemy units as red circles, squares, and stars. Instead of individual names, the units in this game were labelled with types ("you" for the player-controlled character, and "basic", "heavy", or "ranged" for the enemy units). These choices were made to try to maximize the "personification" of the characters in one version of the game while also increasing the social realism of the game by framing the gameplay in a vaguely historical context.

Apart from these visual changes, both variants of the games behaved and played identically, with a small set of deterministic rules deciding how the player-controlled characters could move, and with a naïve rule-based system used to decide how the enemy units would move.

Participants

Thirteen students, all students or members of staff at the University of York, took part in this study. Four participants identified as female, and the remaining nine as male. Each age category except for the age 17 was represented by at least one participant, with four participants being in the age range 26-30. Participants took part in the study voluntarily, and were not compensated in any way.

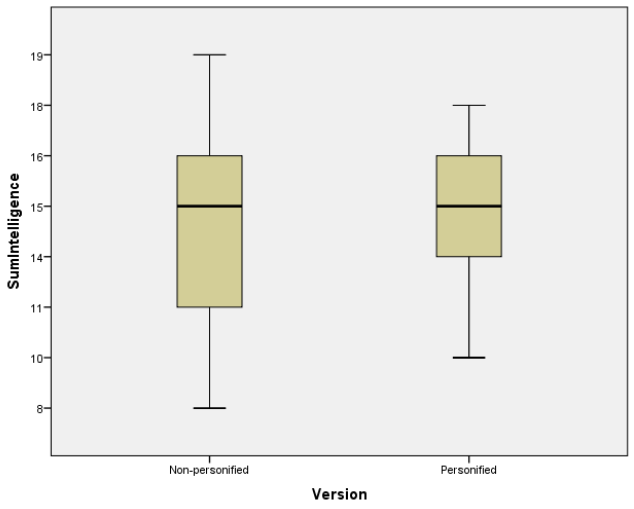

Figure 2: Intelligence measures for the personified and non-personified game variants

Due to the time constraints of this assessment eight of the participants were taken from the Computer Science department, three from the York Centre for Complex Systems Analysis group, and one from the Management department. Discussion of the potential impact of this selection of participants can be found in the section Discussion on the Method Used later in this document.

RESULTS

Each of the participants played for the entire five minute slot, with each participant getting to at least the eleventh level of the game, with a single participant getting as far as the fourteenth and final level although not completing it.

In order to get a single measure approximating how intelligent each participant considered the AI to be a measure was created by summing the values of the answers of questions 10-14, creating a value in the range 5-25. Box plots for these values can be seen in Figure 2.

As seen in the box plots, there is minimal difference of the intelligence measures between the personified and non-personified game variants, with a mean score of 14.5 (σ = 2.66) for the personified variant and 13.6 (σ = 3.73) for the non-personified variant.

Due to the limited number of question used to generate this intelligence metric, we are unable to consider the intelligence measure as parametric. As such, a Mann-Whitney U test is used, giving the value $U = 19.5$ ($p = 0.83$) meaning that there is no significance in the claim that the intelligence metric differed between the variants of the game.

When looking at the answers to the individual questions we are able to some potentially interesting results, such as a higher average difficulty score for the personified variant ($\mu = 3.3$, $\sigma = 0.52$, against $\mu = 2.9$, $\sigma = 0.38$) and a higher score for the frequency in which users felt outsmarted by the AI for the personified variant ($\mu = 3.3$, $\sigma = 0.52$, against $\mu = 2.7$, $\sigma = 1.11$). Further discussion on this is given below.

DISCUSSION

Discussion on Results Achieved

The hypothesis stating that players would consider artificial agents that appear human-like more intelligent than those that were represented abstractly was not supported by the experiment that I performed. Although a minor increase in the mean intelligence score was observed for personified characters, the differences between the observed means was not statistically significant for the samples size used.

This appears to support the results in [4], although it may be the case that limitations upon these experiments are partially responsible for the result achieved. Some discussion on these limitations is given in the "Discussion of the Method Used" section below.

As previously noted, we are able to see some potentially interesting results when examining the scores to individual questions. It is important to realize that given we hadn't begun the experiment with hypotheses around the effect of personified characters on the difficulty or ability of the players to feel outsmarted, we should not consider these as significant results.

These results may, however, be considered as features that justify further experimentation with custom-made experiments in order to see whether they are actual correlations or just peculiarities in the data from this experiment.

Discussion on the Method Used

Due to the constrained nature of this project, certain limitations may have had some impact on the results I have managed to achieve, and may impact the validity of the work performed.

An obvious limitation of this experiment is the limited number of participants used, with 13 participants in a between-participants study being a low number for attempting to get significance in experiments where you wouldn't expect to see a vast difference. By comparison, in [3] 21 participants and within-participants methods were used, meaning that 42 immersion measures were used to assess the impact of priming participants on their immersion in the game. This is not to say, however, that significance could have necessarily been achieved by using more participants, but rather that by using a smaller number of participants that the likelihood of achieving significance is much lower.

The type of participants used in experiments can often have a great impact on the results achieved. For this experiment, many of the participants used were taken from the University of York Computer Science department building, which may mean that the participants used may have a greater understanding of the area of Artificial Intelligence than the public, or than the average game player. Given that I was asking the participants for their opinions of the AI in HexGame you may be able to conclude that the responses may be biased by their knowledge.

While debriefing the participants, several did mention that they thought they perfectly understood how the enemy units were choosing which move to take, and subsequently rated the AI as less intelligent. Whether this is a flaw in the AI used, or rather a problem introduced by asking Computer Science students and staff to rate the intelligence of simple AIs, is unclear, but it could give rise to why a certain proportion of the participants did not consider the games AI as even slightly intelligent.

This may also demonstrate another problem with the experiment as within HexGame players are made to retry levels whenever they fail and because the enemies movements are entirely deterministic, meaning that a player who is made to repeat a level multiple times would be likely to notice that the enemies are performing the same actions. From this players may then abstract from the idea of them being enemies, but rather consider them as pieces in a logical puzzle, reducing the chance that they could see the enemies as intelligent. It therefore may have been more appropriate to use a more complex game that reduces the possibility of the exact same game state being reached multiple times by a single player, or to use a game with non-deterministic AI.

As the game used in this experiment had to be quickly generated to fit in the time-frame required of this experiment, the amount of time put into making sure that the game itself was introducing no biases into the results was less than desired. It is not unlikely that the particular choices of personified and non-personified agents will have contributed to the results achieved, and given that a very limited amount of time was given to creating these graphical styles and justifying the choices made it is currently unclear how big this impact may be. The minor increase of difficulty observed for the personified game variant may, for example, be due to people finding it harder to distinguish between enemy types in the personified game variant than the non-personified game variant.

CONCLUSIONS

In the experiment I have undertaken for this assessment I have been unable to prove that personifying artificial agents in games has any significant impact on players' perceptions of the agents' intelligence.

Further Work

Further work in this area could attempt to address the shortcomings of the experiment. In particular, if the experiment were to be repeated using a different game that has more potential for agents to perform intelligent actions and where greater consideration was given into how to generate multiple representations of the enemy units then we may be able to gain a better understanding of whether there is any difference in how players view agents that are or are not personified.

Further work should also try to ensure that a more representative set of participants are used, and that more participants are used, in order to test the hypothesis in a way that can be generalized more easily.

Any follow-on studies could also look into whether the correlations between personification and difficulty or personification and the feeling of being outsmarted that I have informally observed are based in fact rather than being random correlations in my test data.

REFERENCES

[1] S. Johnson, "Playing to Lose: AI and Civilization," 30 Aug 2010. [Online]. Available: http://www.youtube.com/watch?v=IJcuQQ1eWWI. [Accessed 31 Oct 2014].

[2] M. van Lent and J. Laird, "Developing an Artificial Intelligence Engine," in *Game Developers Conference*, San Jose, 1998.

[3] A. Denisova and P. Cairns, *The Placebo Effect in the Context of Adaptive Artificial Intelligence in Digital Games,* 2014.

[4] T. Koda and P. Maes, "Agents with faces: the effect of personification," in *1996., 5th IEEE International Workshop on Robot and Human Communication*, 1996.

[5] A. McMahan, "Immersion, Engagement and Presence," in *The Video Game Theory Reader*, Routledge, 2003, pp. 67-86.

[6] Magma Fortress, "Magma Fortress: Hoplite," 2014. [Online]. Available: http://www.magmafortress.com/p/hoplite.html. [Accessed 31 Oct 2014].

APPENDIX

Questionnaire presented to participants (with suggested answers removed)

1. Name
2. Sex
3. Age
4. How often do you play digital games? This includes console games, PC games, and games on your mobile
5. When you play digital games how long do you usually play for in a single session?
6. If you play digital games regularly which games have you played in the last week?
7. In general which sort of games do you prefer to play?
8. How enjoyable did you find the game you just played?
9. How difficult did you find the game you just played?
10. How intelligent would you say the enemy units were?
11. To what extent did the enemy units react to your moves?
12. To what extent do you think the enemy units were forming plans?
13. How complex would you say the game's artificial intelligence was?
14. Did you ever feel outsmarted by the enemy moves?

Data recorded by the experimenter

1. Game version used
2. End condition met
3. Level achieved
4. Player deaths