

A Study on Human like Characteristics in Real Time Strategy Games

Johan Hagelbäck and Stefan J. Johansson *Member IEEE*

Abstract—Computer controlled characters (NPCs) are important in any video game to make the game world interesting, give more depth to a game and make the game playable. In almost any game the player has to cooperate with, fight against or interact with NPCs. This is especially true for single-player games but NPCs are also important in most multi-player games. When creating NPCs the developers often strive to create human like characters that behave reasonably intelligent in most cases. We have performed a study aiming to give an idea of the characteristics of human like NPCs in real-time strategy (RTS) games. In the study participants were asked to watch a recording of an RTS game and decide and motivate if the players in the game were controlled by a human player or a computer. We recorded matches where human players played against bots as well as bots playing against other bots. The results were categorized into different groups and they showed that some characteristics, for example simultaneous movement, are perceived as very bot-like and other things such as ability to try different tactics are perceived as humanlike.

I. INTRODUCTION

Games contain a large amount of computer-controlled characters (we will in the rest of the paper use the abbreviation NPCs, non-player characters). They can be opponents the players has to beat, team members the player has to cooperate with, or inhabitants in cities which the player might or might not be able to interact with. The purpose of computer-controlled characters is to populate the game world and give more depth and playability to the game. This is especially important in single-player games where the computer-controlled characters should give the player a challenge but not be overpowered, and make the game fun to play. Because that is why we play games, for fun!

A. Humanlike NPCs

The goal is often to create interesting NPCs by making them more humanlike. It is debatable if a more humanlike NPC is more fun to play against. Yannakakis and Hallam mention in [1] that this does not always has to be the case. Soni and Hingston created humanlike bots in the first-person shooter (FPS) game Unreal Tournament by letting human players train bots how to behave using neural networks. Their experiments showed that the bots trained to behave more humanlike were clearly perceived as more fun (and humanlike) than the coded bots [2]. Freed et.al. has made a survey to identify the significant differences between human and computer play in the RTS game Starcraft. They conclude that more humanlike bots can be valuable for training new

players in a game as well as providing the players with a better testing ground for new tactical and strategic ideas [3]. They also mention that one of the main problems when creating such bots is to understand the significant aspects of how humans play a game.

We are not really at the point where we can create actual intelligence, so we have to focus on what Scott defines as "The Illusion of Intelligence" [4]. The NPCs does not have to be intelligent, as long as they do reasonably intelligent actions under most circumstances. What is reasonably intelligent and humanlike is of course very genre specific. A soldier sneaking through a jungle full of enemies has to take very different decisions than a basketball player in the middle of an exciting match.

B. Description of the study

We have performed a study where participants were asked to observe a replay of a Real-Time Strategy (RTS) game. The replays were created by recording an RTS game in which two players were to create an army and defeat the opponent. The players could either be two computer-controlled (bots), or a bot and a human player. The participants observing the replays were asked to tell if they thought the players were controlled by bots or humans, and motivate their answer. The goal of the experiment was to see what characteristics that are considered typically humanlike. Some might draw a parallel with the famous Turing test, but our aim is not to develop a bot that plays like a human would, but rather try to find what is perceived as human like decisions in an RTS game.

The same experiment setup was used by Livingstone in [5] but with the well known game Pong instead of an RTS game as we use. Livingstone concludes that this experiment setup cannot be called a Turing Test since, for example, there is no interaction between the participants and the AI bot. He also concludes that believability tests can provide interesting insights and ideas for improvements and can be very useful if care is taken when setting up the experiments and analyzing the results.

C. Outline

First, we will describe the environment and how we generated the replays that were used in the study. This is followed by a description of the experimental setup. In Section IV we describe the results and a discussion about them are presented in Section V. We finish by presenting some ideas about future work.

Both authors are with the School of Computing, Blekinge Institute of Technology, Karlskrona, Sweden (phone: +46-455-385831; email: {jhg,sja}@bth.se).

II. GENERATING REPLAYS

In the experiment we used the open-source RTS game engine Spring and the mod Evolution RTS [6]. Evolution RTS is a science fiction mod where the players have access to a wide range of different futuristic tanks and buildings. In the experiments only ground units were available. Players were not allowed to build any aircraft- or hovering units. The ground unit types range from fast, light armored scout tanks to large, very heavy but slow tanks with massive firepower. There are also two different artillery unit types, which can deal heavy damage but have problems hitting moving targets. To defend a base two stationary defensive towers, one light and one heavy, are available. The players also need to gather resources in form of energy (gained from solar collectors and power plants) and metal (gathered from metal extractors placed at specific areas on the map).

Five persons were asked to play one game each against a bot. The persons were given a brief introduction to the game and played a practice game against a bot available in the Spring repository (RAI version 0.601) before the actual logging took place. The recordings started once the second game began and were stopped when one of the players were defeated. The five persons participating in the recordings had varying experience in RTS games.

The bots were implemented using the multi-agent potential field approach described by Hagelbäck and Johansson in [7], [8], [9]. Two different bots were created, called Largebase and Tankrush. The details of each bot are described in Sections II-A and II-B.

The two bots come in three different versions. The differences are in the paces in which the bots perform their actions. In version 1 the bot immediately executes an action once it has decided what to do. In version 2 there is a slight delay between when an action is finished and when a new one can be started. Version 3 has a slightly higher delay than version 2. The delay slows down the bot and can be perceived as having a more humanlike thinking and reaction time. The delay is random between certain intervals. Version 2 has a delay between 4 and 8 seconds, and version 3 a delay between 8 and 15 seconds, see Table I.

TABLE I

THE SIX BOTS USED IN THE EXPERIMENT. THE DELAY IS THE NUMBER OF SECONDS IT WAITS AFTER EXECUTING ONE ACTION, BEFORE IT STARTS EXECUTING THE NEXT.

Bot	delay (in s)		
	0	4–8	8–15
Largebase	L_1	L_2	L_3
Tankrush	T_1	T_2	T_3

Fourteen replay games were recorded for the experiment (see Table II). In five of those a human was playing, and in the other nine different versions of the bots were playing against each other.

TABLE II

THE FOURTEEN REPLAYS USED IN THE EXPERIMENT.

Match	Player 1	Player 2	Winner	Time (min)
M_1	L_1	R_1	L_1	12
M_2	L_1	R_2	L_1	13
M_3	L_1	R_1	L_1	19
M_4	L_2	Mike	Mike	12
M_5	L_3	R_1	L_3	15
M_6	L_2	R_2	L_2	12
M_7	L_1	Pete	Pete	28
M_8	L_3	R_3	L_3	14
M_9	L_1	R_3	L_1	13
M_{10}	L_2	George	George	18
M_{11}	L_3	R_2	L_3	24
M_{12}	R_2	Simon	Simon	14
M_{13}	L_2	R_3	L_2	11
M_{14}	L_1	Mike	Mike	32

A. Largebase

The Largebase bot focuses on building a large base and using heavy units. First it creates ten light scout tanks, which he uses for an early surprise attack. If the opponent is still alive the bot builds a second base in an opposite corner of the map. Both bases are protected by 2-4 light, defensive towers. The bot builds groups of the most heavy tank available (Fatso) with some support from artillery units. Once a group of four heavy tanks and 2 artillery units has been built, the group locates and attacks the enemy. The same group setup is used in the rest of the game.

B. Tankrush

The Tankrush bot focuses on quickly getting some resource income and then build groups of medium heavy tanks supported by artillery. Once a group of 5 tanks and 3 artillery units have been built, the group is sent to attack the enemy. The same tactic is used throughout the game. The bot does not use any defensive towers.

C. The human players

Four different human players participated in the replay games; we call them Mike, Pete, George and Simon. The former three are PhD students, and the latter one is a student at the bachelor level.

Mike as well as George played RTS games around 10 years ago and they consider themselves to be averagely skilled RTS players. Simon is an experienced gamer, but he has not played very much RTS games. Pete is a quite experienced player with an understanding of RTS game AI mechanics.

III. EXPERIMENTAL SETUP

The experiments were carried out at DreamHack Winter 2009, the largest LAN party in the world. We were positioned at the booth of our University, where players who stopped at our booth were asked if they would like to participate in a scientific experiment. Those who agreed were told that they should watch a replay of a RTS game, and what they were to observe and make a decision about. During the game the participants were informed about the type of units

and structures each player was using, and the strengths and shortcomings of the units. Once the participant has made up a decision of if the players were human- or bot-controlled, they were asked to fill in a questionnaire.

The same experiment was carried out at a local LAN party held at a high school close to our University.

A. The Questionnaire

The Questionnaire consisted of a brief instruction on how to fill in the answers, a number of general questions (concerning age, sex and experience of playing RTS games), and a section for each of the two players. Each section contains three word pairs where the participants were to rank how they perceived the player played the game, a checkbox asking if the participant thought the player were a bot or a human, and a free text motivation for their decision (see Fig. 1). The word pairs used are listed in Table III. For each word pair the participants gave a score between 1 (the player is best described by word 1) and 7 (the player is best described by word 2). A score of 4 means that the participant does not think any of the two words describes the player.

Utvärdering av Spel-AI DH 2009

Instruktioner
Fyll i de första frågorna. Vi kommer sedan att visa dig en RTS match mellan två spelare. Ingen, en eller två av spelarna är mänskliga; eventuella resterande spelare är datorstyrda. Du kommer efter matchen att få besvara ett antal frågor om spelarna, så titta noga.

Före matchen:
Min erfarenhet av att spela realtidsstrategispel sedan tidigare är:
OBEFINTLIG MYCKET LÅNG

Jag är:
MAN KVINNA

Min ålder är:
≤ 10 år 11 – 15 år 16 – 20 år 21 – 25 år 26 – 30 år ≥ 31 år

Efter matchen:
Hur skulle du beskriva den röda spelaren?
FÖRUTSÄGBAR OMVÄXLANDE
LÄTTSLAGEN KRÄVANDE
TRÅKIG UNDERHÅLLANDE

Jag tror den röda spelaren är:
EN MÄNNISKA DATORSTYRD

därför att:

Hur skulle du beskriva den gula spelaren?
FÖRUTSÄGBAR OMVÄXLANDE
LÄTTSLAGEN KRÄVANDE
TRÅKIG UNDERHÅLLANDE

Jag tror den gula spelaren är:
EN MÄNNISKA DATORSTYRD

därför att:

Om du vill vara med och tävla, tipsa resultatet i matchen mellan vår bästa bot och Stefan Petersson, lärare på BTH. Matchen spelas live i monter på Lördag kl 19.00. För att kunna vinna något av priserna måste du vara närvarande vid prisutdelningen! Om inte, kommer en ny vinnare utses.

Löppnummer: 146 Spel nummer 2

Fig. 1. The questionnaire used in the experiments.

IV. THE RESULTS

In total 56 persons participated in the experiment. A majority of the participants were between age 16 and 20. Of the participants 52 were male and 4 female.

The results will be presented in detail for each bot individually. The upper part of Table IV–IX describes in what match

TABLE III

A LIST OF THE WORD PAIRS USED IN THE QUESTIONNAIRE.

Word 1	Word 2
Predictable	Varied
Easily beaten	Demanding
Boring	Entertaining

(M_1 – M_{14}) the bot played, what opponent it faced, and the quotient between how many of the participants that judged it to be a human player (H) and a computer bot (C). The last three columns refer to the judgement of the predictability (Pred.), the strength and the level of entertainment (E-mnt) it would provide. Each of these values is an average of the participants' scores along an ordinal scale 1–7.

The lower part of the tables present the written arguments that the participants left to defend their guesses of which players were human and which were bots. These comments have been translated from Swedish. Each separate quote is the translation of the answer of one participant, but not all participants commented on their guesses (so there may be more participants that studied a match than the ones commenting on their classification of the players).

In Table X we show the results of the human players in a way similar to the one for the bots.

A. The Largebase Bots

The results of the Largebase bots L_1 – L_3 is shown in Tables IV–VI.

TABLE IV

RESULTS OF THE MATCHES WITH THE L_1 BOT.

Match	Opponent	H/C?	Pred.	Strength	E-mnt
M_1	R_1	3/2	4.4	5.8	5.6
M_2	R_2	3/0	5.7	4.0	4.0
M_7	<i>Pete</i>	2/1	4.3	4.3	4.0
M_9	R_3	1/2	5.3	4.3	5
M_{14}	<i>Mike</i>	1/1	4.0	2.0	4.0
Average		10/6	4.75	4.44	4.69
Hum.?	"He is thinking strategically and awaits the opponent.", "More aggressive, built. Began light and changed to heavier. Moved its units.", "Just a feeling.", "He is too flexible to be a bot.", "Uses self destruct and scouts with small units. Builds good counters against opponent.", "More active. Spread out fast. Fast attack. Concentrated attacks.", "He changed his tactics very often", "Plays more defensive.", "He moves straight on."				
Comp.?	"It moves all units separately.", "Fast to perform new tasks once a building is finished.", "Did not attack with its full strength.", "Feels as if he controlled most of his units simultaneously, which made it feel like a bot. Also went straight towards the opponent base."				

B. The Tankrush Bots

In Tables VII–IX the participants opinions about the Tankrush bots R_1 – R_3 are shown.

C. Human player results

Table X shows the results of the participants' opinions about the human players.

TABLE V
RESULTS OF THE MATCHES WITH THE L_2 BOT.

Match	Opponent	H/C?	Pred.	Strength	E-mnt
M_3	R_1	6/2	4.1	3.4	3.4
M_4	Mike	3/2	5.4	3.2	4.0
M_6	R_2	0/3	4.3	4.3	3.7
M_{10}	George	0/6	3.8	3.3	3.3
M_{13}	R_3	1/3	3.2	3.5	4
Average		10/16	4.19	3.46	3.62
Hum.?	"Looks like it, the way he plays.", "It feels as if there was a thought behind each move, even though it did not succeed fully.", "In the way he built and moved.", "He/it does not have the time to move its units, since humans cannot do that much simultaneously in strategy games.", "The way the buildings were placed", "Seemed to make human mistakes that a computer never would have done.", "Red is poor at doing lot of things simultaneously.", "Built so little."				
Comp.?	"I personally think it had a quite CPU-like pattern in its game play.", "Calm play. Hazardous unit build planning. Too little 'base' defense line.", "Consequent strategy", "He sent bad units on a good unit to defend a metal extractor. Not very large advancement.", "Even a beginner would have been better.", "It is fast. I would have guessed at high APM.", "Faster, kind of.", "It builds so much"				

TABLE VI
RESULTS OF THE MATCHES WITH THE L_3 BOT.

Match	Opponent	H/C?	Pred.	Strength	E-mnt
M_5	R_1	3/0	6	4.3	3
M_8	R_3	2/1	4.7	6	6
M_{11}	R_2	2/3	4.4	5.2	3.8
Average		7/4	4.91	5.18	4.18
Hum.?	"He was unpredictable, not like a bot.", "He used a lot of dynamic methods that feels more typical to an experienced player than a bot.", "It builds another camp", "Two bases", "He seems to know where the opponent is, even though he has not discovered that much of the map.", ""				
Comp.?	"He did not build in patterns as 2 did. And he controlled several men simultaneously, he built strong defense.", "Builds small, unimproved tanks. Attacks un-smart.", "He was the more eager one and attacked almost immediately."				

D. Human-like vs. bot-like RTS game play

In Figures 2–4 the participants' expressed opinions about the predictability, the strength and the entertainment value of the bots, as well as the players, are shown. Although there seem to be no general conclusions that can be drawn based upon the results, there seems to be a correlation between the predictability of a player, and the strength of it.

The next step is to categorize the comments into different groups. This can of course be done in several ways. We have identified the following: *simultaneous movements, overall strategy, attack tactics, placement of buildings, level of activity, ability to adapt, build order, player strength, map knowledge, intelligence and creativity, and monotony*. There is also a number of very vaguely expressed *feelings* that it is human or computer based. These comments are collected in an *emotional judgement* class. If a participant left a motivation spanning over more than one category, we have filed it as several comments (which is why the sum of experiments does not fully match the number of comments).

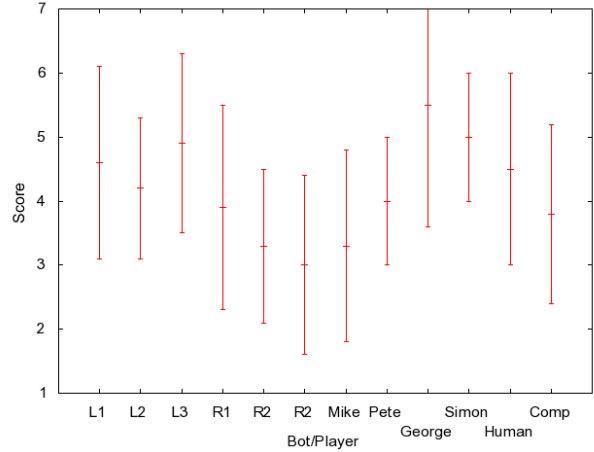


Fig. 2. The predictability of the players according to the participants.

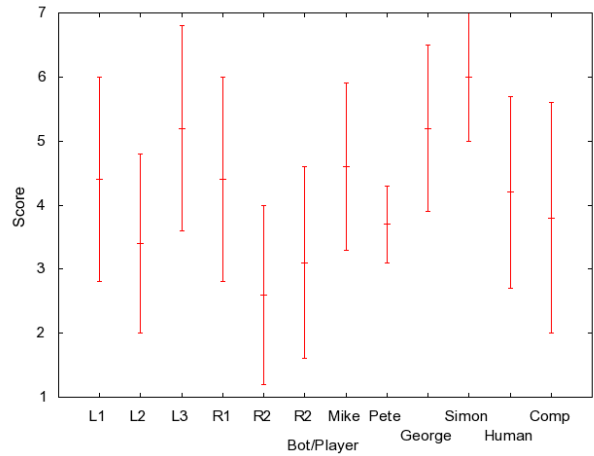


Fig. 3. The strength of the players according to the participants.

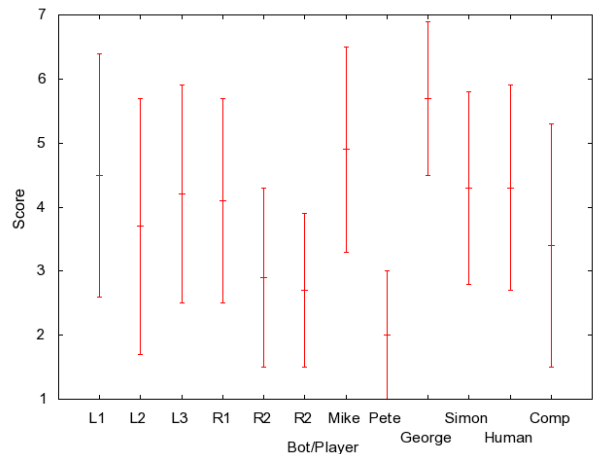


Fig. 4. The entertainment value of the players according to the participants.

TABLE VII
RESULTS OF THE MATCHES WITH THE R_1 BOT.

Match	Opponent	H/C?	Pred.	Strength	E-mnt
M_1	L_1	2/3	3.2	3.8	3.6
M_3	L_2	2/6	4.4	4.9	4.3
M_5	L_3	0/3	3.7	4	4.3
Average		4/12	3.88	4.38	4.06
Hum.?	"Organized attacks.", "He runs more 'tactics'. He tries to take one base at a time."				
Comp.?	"He goes 'straight for it'", "Does not only produce the same tanks. Not very defensive.", "He was predictable.", "Same as 1 (I personally think it had a quite CPU-like pattern in its game play.)", "Stayed and patrolled unnecessary areas.", "All units moved at once.", "The structure of the base and the way the machines moved.", "Always has its troops in movement.", "Use a lot of men at the same time."				

TABLE VIII
RESULTS OF THE MATCHES WITH THE R_2 BOT.

Match	Opponent	H/C?	Pred.	Strength	E-mnt
M_2	L_1	0/3	3.3	1.7	2.7
M_6	L_2	2/1	3.0	1.7	2.3
M_{11}	L_3	3/2	3.2	3.2	2.6
M_{12}	<i>Simon</i>	2/1	3.7	3.7	4.0
Average		7/7	3.29	2.64	2.86
Hum.?	"He did not do very much.", "He built in patterns and without defense towers.", "He built crosswise.", "It upgraded first before it attacked.", "Same thing, they kind of know where the other is from the start. They do not explore and starts going with fairly strengths at once."				
Comp.?	"It attacks the most.", "Very still. Did not move its units. Did not change tactics.", "Too correct built.", "Makes it easy for the opponent.", "Monotonous.", "A slightly better bot that upgrades, but it did not look like a human player."				

1) *Simultaneous movements*: The ability to control many units individually and simultaneously is a typical indicator of a computer player, according to the participants. 9 out of 9 comments also correctly identified bots using this property, compared to 1 out of 3 correct identifications of human players. The two that did not manage to be helped by this, guessed that it was human players because they moved slowly or seemed unable to make things simultaneously.

2) *Overall strategy*: This includes general strategic impressions such as the willingness to attack and to defend itself. In total 17 comments were on the general strategy of which 10 were in favor of a human player, and 7 of them supported a bot. 6 guesses of each sort were correct.

3) *Placement of buildings*: This category includes comments about where the players placed their buildings, often in relation to the buildings already built.

4) *Build planning*: Although seemingly similar to the previous category, this covers the order in which the players seemed to build their units (both buildings and mobile units such as tanks). From the figures we can see that it seems to be a far better indicator for identifying bots, than to identify human players.

5) *Tactics*: The tactics was one of the most successful indicators of whether the player was human or not, and that applies to both types of players. The category consisted mainly of comments on the way it attacked the opponent.

TABLE IX
RESULTS OF THE MATCHES WITH THE R_3 BOT.

Match	Opponent	H/C?	Pred.	Strength	E-mnt
M_8	L_3	1/2	2.0	1.7	2.3
M_9	L_1	1/2	3.3	3.7	3.0
M_{13}	L_2	3/1	3.5	3.8	2.8
Average		5/5	3.00	3.10	2.70
Hum.?	"Felt a bit new to the game. Did not go straight for the opponent. Did not retire when he should.", "Classical opening moves. Focus on collecting resources.", "It plays somewhat slower, but stronger.", "More careful and safe?"				
Comp.?	"It moves all units separately.", "He played generally stiff, sent units at an overwhelming defense without retiring and expanded very little.", "Same as 1 (in the way he built and moved)", "It is very slow and dull.", "It built in a middle of a war."				

TABLE X
RESULTS FOR THE HUMAN PLAYERS.

Match	Player	Opp.	H/C?	Pred.	Str.	E-mnt
M_4	<i>Mike</i>	L_2	2/3	3.2	4.8	5.0
M_7	<i>Pete</i>	L_1	1/2	4.0	3.7	2.0
M_{10}	<i>George</i>	L_2	6/0	5.5	5.2	5.7
M_{12}	<i>Simon</i>	R_2	1/2	5.0	6.0	4.3
M_{14}	<i>Mike</i>	L_1	2/0	3.5	4.0	4.5
<i>Mike</i>	Hum.?	A bit monotonous. Goes directly for the largest tank and thinks little about defense and the tactics of the opponent.", "The yellow attacked directly", "It is slow and do not know what it is doing."				
<i>Mike</i>	Comp.?	"Did exactly what you could expect. No real surprises.", "Built too much."				
<i>Pete</i>	Hum.?	"Forgot units as he was being attacked."				
<i>Pete</i>	Comp.?	"A little bit too passive.", "He was very defensive and did the same thing all the time."				
<i>George</i>	Hum.?	"More impulse.", "The way the units were moved."				
<i>Simon</i>	Hum.?	"Thinks more about defense."				

6) *Level of activity*: If it seemed slow, or fast, the comment on it ended up here. This was in our experiment, a reasonably good indicator of a computer player.

7) *Player strength*: Not very many participants commented on the strength of the player, but for those who did, the guesses were correct.

8) *Game world knowledge*: Behind this indicator is both the ability to seem to know where the opponent is (i.e. a perceived lack of fog of war) as well as the (lack of) ability to perform exploratory tasks. Also this seems to be quite an informative indicator when it comes to make qualified guesses.

9) *Intelligence and creativity*: This category includes comments about the level of intelligence of the player, e.g. referring to "human mistakes that a computer never would make". This indicator was not very successful for the participants and most of the guesses failed.

10) *Ability to adapt*: The ability to adapt was according to the participants an ability that human players had. All of them were in this case wrong in their guesses.

11) *Monotony*: Monotony refers to the identification of repetitive actions (or lack thereof).

TABLE XI
CLASSIFICATION ACCURACY OF THE DIFFERENT TYPES OF COMMENTS.

Comment category	Total	Correct guesses		Incorr. guesses	
		Human	Bot	Human	Bot
Simult. movements	12	1	9	2	0
General Strategy	17	6	6	4	1
Placement of buildings	4	2	2	0	0
Build planning	17	4	8	4	1
Tactics	9	4	4	1	0
Level of activity	10	2	6	1	1
Player strength	4	2	2	0	0
Game world knowledge	7	3	3	1	0
Intelligence/creativity	6	2	0	3	1
Ability to adapt	3	0	0	3	0
Monotony	7	1	4	1	1
Emotional judgement	7	2	3	2	0
Sum of comments	103	29	47	22	5
No comment given	33	5	16	9	3

12) *Emotional judgement*: In this category, we included vaguely expressed statements such as: "CPU-like pattern in its play" (bot), or "Looks like it the way he plays" (human).

13) *No comment*: A quite substantial part of the participants did not leave any (or just a partial) motivation to their guesses. In all, 33 guesses were left without comments. 21 of them were correct, and 12 incorrect.

V. DISCUSSION

The aim of the study was to see if we could find some characteristics of human like behavior in RTS games. The aim was not to create a bot that behaved like a human, neither to show that human players were most often perceived as humans in the replays (even though we provided these numbers in the paper).

The results, when categorized, showed some interesting things. *The movement of the units* is one thing that differs between bots and human players. Bots are able to simultaneously move different units in different directions, which is not possible for the average human player. This could be prevented if bots grouped units in squads and let all units in a squad move in the same direction, and not let several squads issue new move orders at the same time.

Properties like *intelligence*, *creativity* and *adaptability* were however not a clear way to differ between bots and humans. Although the majority of participants thought these characteristics were humanlike, some thought that it instead was a sign of a well-designed bot. Some people thought a bot shall react to the opponent and try different unit setups, and that most bots are implemented in that way.

The *Level of activity* and the *player strength* were reasonably good indicators. A player that does too many things will be perceived as a bot, while players with low activity are often seen as inexperienced humans. A player that wins his or her match is often seen as a good human player and rarely a bot. Note that we did not allow the bots to cheat in any way (e.g. by removing fog of war or let their units get more hit points than the human player's units).

General strategy and *build planning* were not very good indicators. A player that did something clever was as often

perceived as a well-implemented bot as an experienced human player. The same goes for bad tactics; it could be a sign of a bad bot or an indication of an inexperienced human. *Tactics* of attacking or defending units and physical placement of buildings were better indicators, but there were few comments in these categories.

Knowledge about the game world also seems to be a pretty good indicator. Bots were keen to go directly to the most logical spot for the opponent player to be located (i.e. the opposite of the map) while players did more scouting. This was noticed and commented by some of the participants.

A. On the method

If we look back; what could we have improved method wise? Although this series of experiments as such is closed, we would like to share some thoughts on this.

First, the variation of the matches could be larger. We used two bots in three different versions, but none of them were tuned to be strong enough to beat any of the human participants. Given that, we cannot say that much about how a strong bot would be perceived by the participants of the experiment. We also restricted the bots to the type of bots that we have implemented ourselves. Maybe a more traditional path-finding-based bot would have been easier to detect for the participants?

Second, due to a random pick of matches (the participants were given one of the matches at random), there were quite large differences between the most frequently watched game, and the least. This in turn makes it harder to analyze the results and draw general conclusions about the different bots.

Third, we should also have had matches with two human players. For technical reasons, this turned out to be hard to achieve with the setup we used, but we think that it may contribute to the results to also have two human players. A majority of the participants guessed at one human and one bot (even when it was two bots), so for the sake of completeness, we should also have had two human players.

VI. CONCLUSIONS

There are a few indicators that we can conclude are associated with human like RTS game play. The micro-management of units, the placement of buildings, tactics, player strength, and game world knowledge are all indicators that the participants used successfully to argue for a correct guess. Intelligence, creativity and ability to adapt are however less successful signs (in our experiment).

We must also point out that the number of participants in our experiments was relatively small and a larger sample would give us more knowledge about the categories with less strong indications.

VII. FUTURE WORK

We believe our findings can be used to design future experiments that can give us even better insight in human characteristics in RTS games. We can learn from our results and create adaptive bots that are not allowed to simultaneously move units in different directions and that does

not cheat regarding game world knowledge. If the same or similar experiments would be conducted with the re-designed bots it could give us even more insight in human-versus bot-like play for RTS games. A good insight in human game play can be important when creating good AI players. As Yannakakis and Hallam points out, humanlike bots does not always have to be fun to play against, but we believe it is a good start to be aware of humanlike characteristics when designing them [1]. Another issue worth further investigations is to find out how and why there seems to be a connection between the predictability of a player, and the strength of it.

ACKNOWLEDGEMENTS

The authors would like to thank the Games and Interactive Systems Laboratory (GISL) research group for their feedback on early drafts of this paper, all the participants of the experiments at DreamHack, and of course the School of Computing at Blekinge Institute of Technology for the support with travels and equipment for the experiments.

REFERENCES

- [1] G. N. Yannakakis and J. Hallam, "Towards Optimizing Entertainment in Computer Games," *Applied Artificial Intelligence*, pp. 21:933–971, 2007.
- [2] B. Soni and P. Hingston, "Bots trained to play like a human are more fun," in *Proceedings of the 2008 International Joint Conference on Neural Networks (IJCNN 2008)*, 2008.
- [3] M. Freed, T. Bear, H. Goldman, G. Hyatt, P. Reber, and J. Tauber, "Towards more human-like computer opponents," *AAAI Technical Report SS-00-02.*, 2007.
- [4] B. Scott, "The Illusion of Intelligence," in *AI Game Programming Wisdom*. Charles River Media, 2002.
- [5] D. Livingstone, "Turing's test and believable ai in games," *Computers in Entertainment (CIE)*, 2006.
- [6] "The spring project," <http://springrts.com/> URL last visited on 2010-02-04., 2010.
- [7] J. Hagelbäck and S. J. Johansson, "Using multi-agent potential fields in real-time strategy games," in *Proceedings of the Seventh International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, L. Padgham and D. Parkes, Eds., 2008.
- [8] —, "The rise of potential fields in real time strategy bots," in *Proceedings of Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 2008.
- [9] J. Hagelbäck and S. J. Johansson, "A Multi-agent Potential Field based bot for a Full RTS Game Scenario," in *Proceedings of Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 2009.