Immersion level and bot player identification in a multiplayer online game: The *World of Warships* case study

Paweł Łupkowski, Violetta Krajewska
Adam Mickiewicz University in Poznań
pawel.lupkowski@amu.edu.pl | ORCID: 0000–0002–5335–2988
krajewska.violetta@gmail.com | ORCID 0000-0001-8296-3385

Abstract: In this paper we present the results of an experimental study of bot identification in a multiplayer online game. Our game of choice for the study was World of Warships. The tested group consisted of 30 subjects (15 experienced players and 15 players without significant experience in this domain). Subjects played the game against bots or against human players. The main hypothesis for the study was that the more immersed a player was, the less accurate s/he will identify the opposing players (as human players or as bots). On the basis of the results, this hypothesis cannot be confirmed.

Keywords: immersion, multiplayer online games, bot player identification, unsuspecting Turing Test

Homo Ludens 1(11)/2018 | ISSN 2080–4555 | © Polskie Towarzystwo Badania Gier 2018
DOI: 10.14746/hl.2018.11.8 | received 31.12.2017 | revised 14.05.2018 | accepted 27.08.2018

1. Introduction

The motivation for this study comes from a somehow unexpected field, namely the Turing Test (hereafter TT) debates. In his seminal paper, Alan Turing (1950) proposed a test for machines. A machine will pass the test when it is capable of having convincing, human-like tele-typed conversation with a human judge (the parties of the test cannot see or hear each other). The proposed test was widely discussed and analysed within many disciplines. One of the main concerns when it comes to TT is the role of a judge. The argument is that the results will be biased as the judge knows that s/he will have a dialogue with a machine (see e.g. Block, 1995 or an overview in Łupkowski, 2011). One of the most interesting propositions on how to resolve this issue was proposed in a short paper by Michael Mauldin (1994) and further discussed by the same author (2009). Mauldin used the *TinyMud* game (a text-based multiplayer RPG game) and introduced a bot (named *ChatterBot*) into the game. He observed that the bot was often taken for a human player. As Mauldin (1994) writes: "The ChatterBot succeeds in the TinyMud world because it is an *unsuspecting Turing Test*, meaning that the players assume everyone else playing is a person, and will give the ChatterBot the benefit of the doubt until it makes a major gaffe" (p. 17).

In our opinion the described idea reaches far beyond TT discussions, as it has direct and practical implications for the design of multiplayer online games. Many such games use bot players in order to make the game more interesting and playful. Moreover, for some multiplayer online games bots are simply necessary to ensure that playing the game will be possible when the number of human players is too small. In our experiment we wanted to check Mauldin's proposal in the context of a modern multiplayer online game, which is *World of Warships* (WoWS) by Wargaming (<http://wordlofwarships.eu/>). In WoWS players can play against teams consisting of human players or against teams consisting entirely of bots, and thus it offers a convenient tool for running Mauldin's unsuspecting Turing Test.

Inspired by TT discussions concerning judges for the test (see e.g. Block, 1995, Loebner, 2009, or Garner, 2009), we decided to check whether bot identification would differ for experienced players and for people who do

not play games or play them only casually. We also employed the concept of immersion, which often appears in game research. After Jennett et al. (2008, p. 643) we understand immersion as "the specific, psychological experience of engaging with a computer game".[1] As we decided to use a game with advanced 3D graphics and well-designed sound effects, we wanted to see whether the immersion level would have its effect on the identification of opposing players in WoWS.

The paper is structured as follows. In the first section we present our methods and procedure. We introduce the research hypotheses and discuss the choice of a game for the experiment. We also describe the research group. The second section covers our results. In the last section we discuss the results and present some ideas for future studies.

2. Methods and procedure

Our research hypotheses for the presented study were the following:

1. There will be a difference in the declared immersion level between the group of experienced players and the group of inexperienced players.
2. The more immersed a player is, the less accurate s/he will identify the opposing players (as human players or as bots).

We find the first hypothesis intuitive. One may expect that when an experienced player is confronted with a task of identifying a bot in a new game, s/he will perform better than an inexperienced one. The reason for this is that the experienced player has encountered bot players previously and has developed certain criteria for bot recognition which may be applied to the new situation.

As for the second hypothesis, we rely on the immersion characteristics provided by Jennett et al. (2008, p. 643–644). They point out that the immersion state may be recognised by the loss of the time flow, the loss of the awareness of the external world, and the engagement resulting in a "being in a game" state. What is more, researchers (Jennett et al., 2008;

[1] For further discussion concerning immersion in game studies see e.g. Calleja, 2007 (for an interesting model for describing and analysing the players' involvement in digital games) or Ermi and Mäyrä, 2005 (for a general model for the gameplay experience).

Brown and Cairns, 2008) report that the immersion state is correlated with the emotional engagement state. One may expect that the higher the immersion level for a player is, the more 'credit' (in Mauldin's sense) s/he will give to the other players in the game (as long as they will be coherent with the game environment).

2.1. The game

In our experiment we have decided to use *World of Warships*. It is an online multiplayer game where two teams of warships battle against each other. Each player is steering one ship in a given battle. The aim is to sink all the enemy ships or to take over special areas on the battlefield (bases). Teams are constituted of randomly chosen players, so the team's strategy has to be established on the fly during the match.

WoWS is suitable for our study for several reasons. First of all, it is advanced with respect to graphics and to sound effects. Controlling the ship is very intuitive as the standard keyboard keys (W, S, A, D) are used to move it, and mouse is used to aim and fire guns. The player can have a first-person view of the ship or change it to a top view (see Figure 1 for game screenshots). Also, the pace of the game is suitable for our research purposes, since it is not very fast due to the nature of a ship battle (in contrast to e.g. *World of Tanks* or typical FPS games, like *CS:GO*). Warships are also durable, which allows for a longer game experience even for inexperienced players. What is more, the game offers a chat for players and the "Quick commands" menu.

WoWS may be played in two modes, against bot players or against human players (the so-called "Cooperative game" and "Random game"). The player's team members are always human players. This makes WoWS suitable for checking Mauldin's unsuspecting Turing Test assumptions.

It is also worth mentioning that despite the fact that WoWS is a game where teams fight against each other, the game does not depict violence against human beings. The focus is on warships as agents in the game. The PEGI rating of the game (see <www.pegi.info>) is that it is suitable for 7-year-olds and that the violence in the game is not explicit. A similar opinion about the violence level of WoWS (1 of 5 possible points) may be found in the review for Common Sense Media <www.commonsensemedia.org>: "Although it's a war game, no blood, bodies shown – only ships catching fire, sinking".

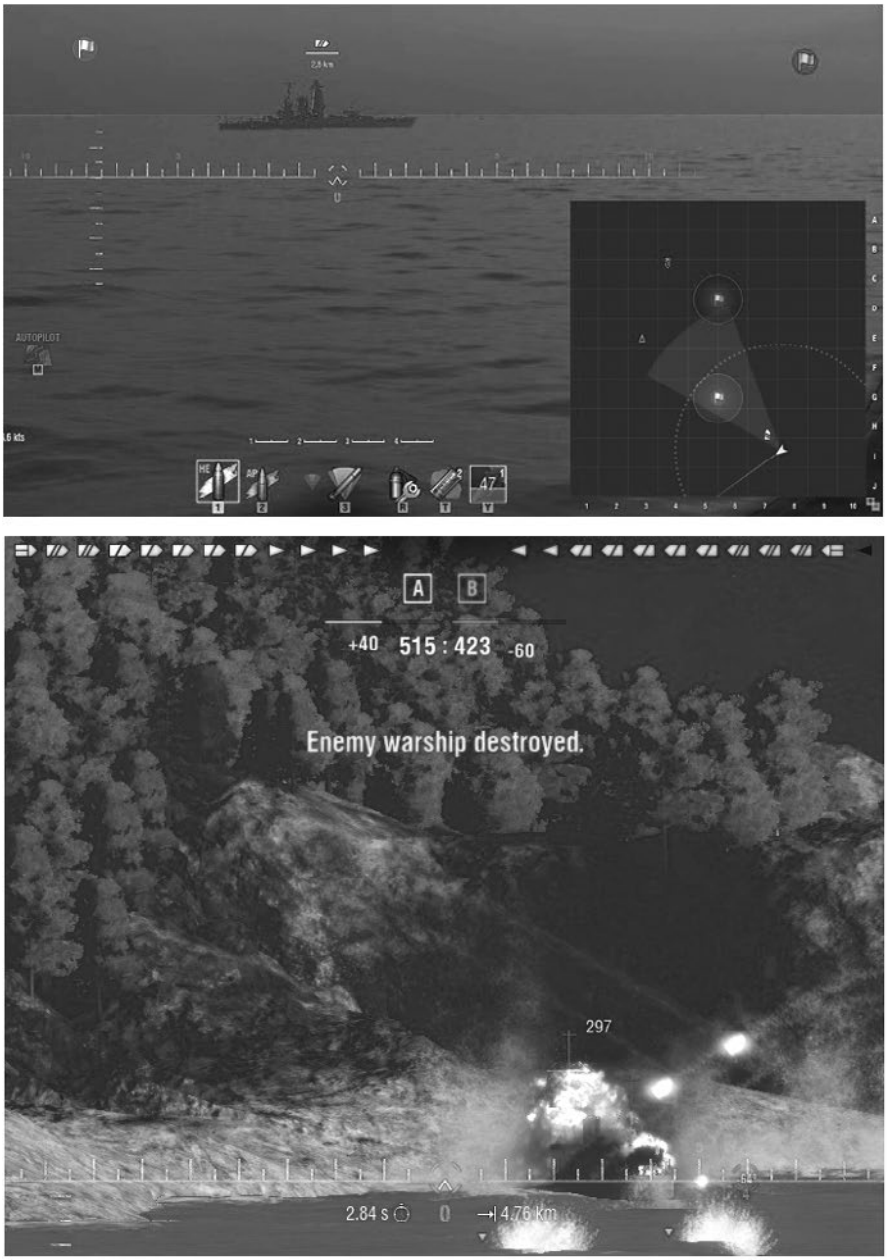

Figure 1. World of Warships *game screenshots*. Source: <http://wiki.wargaming.net>

2.2. Immersion Questionnaire and interview questions

In order to measure the declared immersion level, we applied the *Immersion Questionnaire* (Jenett et al., 2008). As the research was conducted in Polish, we used the questionnaire in its Polish adaptation presented by Strojny and Strojna (2014). The questionnaire was supplemented with one additional question concerning the identification of the opposing players, namely: "Against whom were you playing? A: bots, B: humans".

At the end of the questionnaire, questions about the subjects' age, gender, and game experience (the favoured game genre and the average time spent playing games) were asked. The subjects filled out the questionnaire in an electronic form (Google Forms were used for the presentation and data collection) on the same computer on which they had been playing the game.

After the questionnaire, the experimenter conducted a short interview with the subjects. It consisted of the following questions (here translated into English):

1. Were you comfortable during the game? Maybe something was causing a discomfort? (The intuition behind this question was to identify the potential factors which may influence immersion, like e.g. the experimental setting for the game.)
2. Was controlling the ship troublesome in any way?
3. Was the game interesting for you?
4. Do you think that if you had played a different game, your immersion would have been different?
5. How would you evaluate your interaction with other players in your team?
6. On what elements were you focusing during the game? Did you focus more on your own ship (approaching the enemy from the right distance, not getting damaged), or perhaps on other ships from your team, or on the communication with other players?
7. Against whom were you playing? Why do you think that?

2.3. Procedure

The experiment was conducted in the Reasoning Research Group Laboratory <http://reasoning.edu.pl> at the Institute of Psychology, Adam Mickiewicz University in Poznań. The experiment was conducted by the

second author. For the experiment, 30 subjects who had not played the WoWS game before were recruited. Each subject participated in the study separately. All participants were volunteers, and they did not receive any compensation for the participation.

The experimental setting was the following. Each time only one subject and the experimenter were present in the laboratory. All the subjects were playing on a laptop with a 15.5" screen. An external loudspeaker was used for better sound quality. The subjects were using a laptop keyboard and an external wireless mouse as controls. The game mode ('cooperative' vs 'random') was alternated from participant to participant (with some minor exceptions resulting from the lack of human players in the mornings, because then the cooperative mode could not be used).

The experiment consisted of two main parts: (1) WoWS gameplay and (2) filling the questionnaire followed by a short interview.

Part 1. The WoWS gameplay part covered the training session and the main game. Before the training session each subject read the instruction for the experiment and for the game (i.e. a short introduction to the game rules prepared for the needs of our study; this instruction was at the subjects' disposal for the whole experiment). All subjects were informed about the anonymity of the research and told that they could resign at any moment without providing reasons. The experimenter additionally explained the game rules before the game and answered questions if there were any. Before playing, each participant adjusted the settings for her/his preferences (monitor tilt angle, left- or right-handed mouse settings, volume level, intensity of light in the room). During the training session the experimenter assisted the subject and offered explanations concerning the game elements. It was the subject who decided that s/he was ready to end the training phase. The main game usually took 10 minutes (but not fewer than 8). The subjects managed to play one battle during this time (or sometimes two battles, in the case of experienced players). All subjects used the *Orlan* cruiser in the training session. For the main game the *Hashidate* cruiser was used. When the ship was sunk in the main game, another cruiser of the same class and level was used as a replacement. One subject always played one variant of the game: the cooperative one (against bot players) or the random one (against human players).

Part 2. As mentioned above, the questionnaire was administered with the use of Google Forms and presented on the same laptop that the subjects used to play the game. Afterwards the aforementioned interview was conducted by the experimenter without audio recording – all the answers were noted down by the researcher.

2.4. Subjects

The research group consisted of 30 subjects (12 men, 18 women) aged 19–38 years. We have applied purposive sampling as the research group consisted of 15 subjects who did not play or played only casually and 15 subjects who may be described as experienced players (playing games at least a few times in a week).

The questionnaire contained a question about the frequency of playing games. The numbers of answers provided to this question are presented in Table 1.

Table 1. Research group summary. Answers for the question: "How often do you play games?"

How often do you play games?	Number of answers
I do not play games at all	10
I play games less than once a month	2
I play once a month	3
I play a few times in a week	9
I play every day	6

Basing on the answers to this question, we divided our subjects into two groups: the group of experienced players, who played a few times a week or every day (hereafter referred to as group A), and the group of casual players (group B).

Group A. The subjects from this group prefer role-playing games (8 subjects pointed at this type of games). They prefer multiplayer online games as a form of play (8 subjects).

Group B. In this group we had 10 subjects who declared that they did not play games, and 5 who declared that they played once a month or less than once a month. Out of the latter 5, four pointed simulation games

as their favourite type of games (mentioning *The Sims*). As for the form of play, they prefer single-player games.

3. Results

IBM SPSS 22 was used for data analysis. Regarding the first hypothesis, for the group comparison we used the non-parametric Mann-Withney test for independent groups. Regarding the second hypothesis, we employed the chi-square test and the Yule phi coefficient.

3.1. Hypothesis 1

The *Immersion Questionnaire* reliability as expressed by Cronbach's alpha is satisfactory – 0.871 [Strojny and Strojna (2014) report alpha = 0.938 for their research].

The maximal score in the *Immersion Questionnaire* was 135. For each question a subject obtained between 1 and 5 points (there were 5 questions with reversed scores). The minimal score for the research group was 54 and the maximal one was 122. The mean for the group was 96.87 (SD = 15.61). Ten subjects achieved scores below the mean result. The dominant values were 94, 101, 105 and 112. The median for the group was 100. The summary of these results is presented in Table 2.

Table 2. Results for the Immersion Questionnaire

Group	M	SD	Min	Max	Median
All	96.87	15.61	54	122	98
Group A	95.53	14.23	71	122	98
Group B	98.20	17.24	54	119	101

Our first hypothesis was that experienced players (group A) and casual players (group B) would differ with respect to the immersion level. We also predicted that group A should have a higher immersion level than group B.

Group A. The minimal immersion score for the group was 71 points and the maximal one was 122 (which was the highest score for the whole

research group). The mean score for the group was 95.53 – six subjects from group A scored below this level.

Group B. The minimal immersion score for the group was 54 (which was the minimum value for the entire research group) and the maximal score was 119. The mean value for this group was 98.20 – five subjects from group B scored below this level.

As it may be observed, the mean immersion level for the group A was lower than for the group B. This tendency is not in line with our predictions. What is more, differences in immersion levels between groups A and B were not statistically significant, as shown by the results of the U Mann-Withney test ($U = 89.5$, $p = 0.34$). In conclusion, the first hypothesis was not confirmed. We cannot say that the group of experienced players differed with respect to the immersion level from the group of casual players.

3.2. Hypothesis 2

Our second hypothesis concerned the identification of opposing players as bots or as human players. Our prediction was that the experienced players would be correct more often than the casual ones.

In group A the subjects played 8 times against bots and 7 times against human players. In group B the subjects played 7 times against bots and 8 times against human players. The summary of the results concerning the correctness of the identification process is presented in Table 3.

Table 3. The correctness of the opposite players' identification

Group	Correct	Incorrect
Group A	73.33%	26.67%
Group B	53.33%	46.67%

The results of testing the difference between groups A and B are the following: Yule phi = 0.89, $p = 0.534$. The significance level does not allow us to confirm our second hypothesis.

3.3. Qualitative data

As already mentioned, the study also covered an interview with the subjects concerning the potential factors which might influence the game

experience. As for these factors, it is interesting that they were mainly pointed out by subjects from group A. Moreover, the type of these factors differed for groups A and B. Subjects from group A were focused mainly on the technical settings available in the experiment. For example, they pointed the following factors as negative influences on their gameplay experience (here and below we present the quotations translated into English; the whole study was conducted in Polish): "lack of my own equipment, lack of the mechanical keyboard"; "mouse – bigger would be better"; "lack of headphones, the screen is not big enough, the light is to bright". As for the game itself, subjects from group A claimed that the logging time before the battle was too long and the game itself took too little time. As to the group B, the subjects focused more on the game genre, as expressed by one of the subjects: "not my type of game" (which is in line with the characteristics of this group – see the section *Subjects*).

As for the question concerning game controls, subjects from group A declared that it was intuitive and that the controls were consistent with other games of the same type. Subjects from group B reported more issues with respect to the controls. What appeared to be problematic was that the keyboard and the mouse had to be used to control the ship. Also, the pace of the game was troubling for some participants, as they had to focus on steering the ship and fighting the enemy at the same time. What is more, it was pointed out that the first-person perspective might be difficult as it was hard to grasp the wider perspective of the battlefield ("I have been focusing on the target but I was afraid that I would crash into something"). Another interesting aspect noted by a subject from group B was the responsiveness of the ship. In WoWS ships react with a certain delay to simulate the real process of maneuvering such a large object on the water. However, for our subjects this was disrupting, as the ship was "ponderous". At this point it is worth reminding that the subjects decided themselves when they were finished with the training session (which was aimed at learning and practicing how to control the ship in the game).

The answers to the question about whether the game was interesting also differed in the two groups. Subjects from group A focused on group interaction possibilities, different available strategies, and realistic game physics as positive aspects of WoWS. It is worth stressing that none of these aspects were mentioned by subjects from group B. The latter

focused on the aesthetics of the game (advanced graphics and sound effects). In both groups the subjects addressed the issues of favoured games and game types. They stated e.g. that the game was interesting "because it is a first person shooter, but I like to play something different, like RPG" (group A), or that "it is something different than the usual, I would never try it by my own" (group B).

This issue was also present when the subjects answered the question about the predicted immersion level for a different game. Most of them agreed that changing the game would influence their immersion level. The subjects noted that the factors which – in their opinion – influenced the immersion level for a game were the following: communication and feeling as part of a group, simple task and rules (a player should not be forced into thinking too much – however, there were subjects who named a well-designed plot engaging a player as an important factor), the consistency of game elements, which should not remind the players that they were in a game ("E.g. in *Assassin's Creed* icons appear, they are clearly reminding me that I am playing a game"), and the first-person perspective.

Communication with other players was named as an important factor with regard to immersion. However, when we analyse answers given to the question addressing this issue for the WoWS game in our experiment, it turns out that the subjects declared that they were mainly focused on controlling the ship and shooting the enemy: "No communication, I was just shooting", "More focus on steering", "I only read some chats, but I did not post anything".

As for the question *On what elements were you focusing during the game?*, subjects from both groups declared that their focus was on aiming and firing. What is important, they were mainly describing their actions as individuals, not referring to a group strategy for the whole team: "Bravely, to find an enemy and shoot – attack, and then not to get hit"; "On what I saw, and then what I heard. To get close and then not to get hit". There were players, however, who addressed the team strategy, as in the following example from group A: "I was observing the map and checking what other team members were doing. I was adjusting my actions to the team, I wanted to learn something from them as I was playing for the first time".

The last interview question addressed the issue of the identification of opposite players and the criteria used for this identification. Here

answers in groups A and B were different. Subjects from group B mainly referred to "feelings" and "intuitions" when explaining their choices. As for group A, the subjects said that humans' behaviours in the game were rather "chaotic" and "not organised"; consequently, when the opposing team used a sophisticated strategy, they probably were bots. However, several observations were made that a simple analysis of the strategy was not enough here, as in this statement: "But the team was the same for both, training and the main game? Because when I think about this, I feel that it was not entirely so. In the first game everything was rather coherent because we wanted to capture the base... But the second game, well, that was chaotic". Interestingly, subjects from group A showed much more self-confidence when answering this question. They were often surprised when informed by the experimenter that they had made a wrong identification choice (the interview was the last element of the study).

4. Discussion and Summary

Our two research hypotheses were not confirmed. No significant differences between the group of experienced players and the group of casual players were observed with respect to the immersion level and to the correctness of the identification of the opposite players as bots or humans.

What is interesting (despite the lack of significant difference), we observe that the mean result for the *Immersion Questionnaire* for group A is lower than for group B. One possible explanation for this fact may be that the experienced players were not able to achieve the first barrier for immersion as described by Brown and Cairns (2004), i.e. the engagement barrier. As it was often mentioned in the interview, experienced players lacked their own gaming equipment. They were also pointing out that WoWS was not their favourite game and it was not very interesting for them. This may be noticed in the answers to the interview question about other games and the expected immersion level – almost all the subjects agreed that for their games of choice immersion would be different (often they expected a stronger effect).

In this context it should also be mentioned that the experimental setting itself may be a factor influencing immersion. In the interview the

subjects often addressed the issue of the time limit for the main game, which was evaluated as too short.

As for the second hypothesis, we may search for potential explanations in the interview data. It seems that in group A as well as in group B subjects were mainly relying on their intuitions and ideas on how bots should behave in the game. This is especially visible in the interview, as some subjects took the chaotic behaviours of the opposing team as a sign that it consisted of human players, and at the same time other subjects took it as a premise to conclude that this had to be a bot team.

One of the other possible factors which possibly contributed to the observed result may be the experimental setting again. We recruited subjects who had never played WoWS before the experiment. Perhaps – despite the training session – the subjects in both groups were still focused so much on their own perspective in the game (to approach the enemy and to fire a round) that they were not able to analyse the broader context of a battlefield and the opponents' moves and communications. This is indicated by the interview data, since subjects reported that they had mainly focused on steering the ship, and they had not been communicating with other players. The time restriction for the main game was dictated mainly by the laboratory conditions of the experimental setting and by the fact that each subject was tested separately. We wanted to make the whole procedure short enough for the subject and the experimenter to cope with.

We should also mention that the lack of significant differences for both hypotheses may also be the result of a small research group. The group size was partially the result of the purposive sampling of subjects. Especially for the second hypothesis the observed tendency is promising. Experienced players (group A) were better at recognising the opposing players than subjects from group B. Group A subjects correctly identified all 7 cases where they had been playing against human players and 4 cases where they had been playing against bots. The four observed mistakes were the cases where bots were wrongly taken for human players. As for the subjects from group B, they were more often mistaken when playing against human players (in five cases human players were identified as bots, and in three cases bots were identified as human players). In our opinion, recruiting subjects who were not acknowledged with WoWS was a well justified step for our research. However, future research within the

proposed experimental schema should include longer playing times for the main game and perhaps a training session ending with a short test checking the steering skills. Such a study should also have a larger scale, employing more subjects.

5. References

Block, N. (1995). The mind as the software of the brain. W: E. Smith, D. Osherson, (red.), *An Invitation to Cognitive Science – Thinking* (pp. 377–425). London: The MIT Press.

Brown, E., Cairns, P. (2004). A grounded investigation of game immersion. CHI '04 Extended Abstracts on Human Factors in Computing Systems (pp. 1297–1300).

Calleja, G. (September 2007). Revising Immersion: A Conceptual Model for the Analysis of Digital Game Involvement. In *Situated Play, Proceedings of DiGRA 2007 Conference* (pp. 83–90).

Ermi, L., & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion. In S. de Castell, J. Jenson (eds.), *Worlds in play: International perspectives on digital games research* (pp. 37–53). New York, Bern, Berlin, Bruxelles, Frankfurt am Main, Oxford, Wien: Peter Lang.

Garner, R. (2009). The Turing Hub as a Standard for Turing Test Interfaces. In R. Epstein, G. Roberts, and G. Beber (eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (pp. 319–324). Springer Publishing Company.

Jennett, C., Cox, A., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies, 66*(9), 641–661.

Loebner, H. (2009). How to Hold a Turing Test Contest. In R. Epstein, G. Roberts, G. Beber (eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (pp. 173–180). Springer Publishing Company.

Łupkowski, P. (2011), A Formal Approach to Exploring the Interrogator's Perspective in the Turing Test. *Logic and Logical Philosophy, 20*(1–2), 139–158.

Mauldin, M. L. (1994), Chatterbots, TinyMuds, and the Turing test: Entering the Loebner Prize Competition. In *Proceedings of the Twelfth National Conference on Artificial Intelligence* (Vol. 1), AAAI '94, American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 16–21.

Mauldin, M. L. (2009). Going undercover: Passing as human; artificial interest: A step on the road to AI. In G. B. R. Epstein, G. Roberts (eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (pp. 413–430). Springer Publishing Company.

Strojny, P., Strojna, A. (2014). Kwestionariusz immersji – polska adaptacja i empiryczna weryfikacja narzędzia. *Homo Ludens*, 1(6), 187–198.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 443–455.

Paweł Łupkowski, PhD – Assistant Professor at the Department of Logic and Cognitive Science, Institute of Psychology, Adam Mickiewicz University in Poznań (Zakład Logiki i Kognitywistyki, Instytut Psychologii, Uniwersytet im. Adama Mickiewicza w Poznaniu). Cognitive Science Curriculum Manager at the Institute of Psychology. Co-founder of the Reasoning Research Group <http://reasoning.edu.pl>

Violetta Krajewska, MA – Reasoning Research Group, Institute of Psychology, Adam Mickiewicz University, Poznan.

Poziom immersji a identyfikowanie botów w wieloosobowej grze online — studium przypadku gry World of Warships

Abstrakt: W artykule przedstawiamy wyniki badania poświęconego identyfikacji graczy-botów w wieloosobowej grze online. Badanie zostało przeprowadzone z wykorzystaniem gry *World of Warships*. Grupa badana składała się z 30 osób (15 doświadczonych graczy oraz 15 osób z niewielkim doświadczeniem w grach). Badani grali w grę przeciwko botom lub ludziom. Główną hipotezą badawczą przyjętą przez nas na potrzeby badania było przypuszczenie, że im wyższy poziom immersji osiągnie gracz, tym mniej trafnie będzie rozpoznawał graczy, przeciwko którym gra (jako ludzi lub boty). Hipoteza ta nie znajduje potwierdzenia w wynikach badania.

Słowa kluczowe: immersja, wieloosobowe gry online, identyfikacja botów, unsuspecting Turing Test