

## ***The Cooperative Board Game THREE. A Test Field for Experimenting with Moral Dilemmas of Human-Robot Interaction***



Paweł Łupkowski

(Adam Mickiewicz University in Poznań; Pawel.Lupkowski@amu.edu.pl)  
ORCID: 0000-0002-5335-2988

Aleksandra Wasielewska

(Adam Mickiewicz University in Poznań; alewas1@st.amu.edu.pl)

### **Introduction**

The aim of this paper is to present the cooperative board game THREE as a tool which provides a test field for experimenting with the moral dilemmas arising from human-robot interaction.

The game is inspired by the Three Laws of Robotics (hence the name) invented by the science fiction writer Isaac Asimov and introduced in his 1942 short story "Runaround". The Laws provide the framework for all the decisions and actions taken by robots. Despite the fact that the Laws were devised for the needs of science fiction, they quickly got into the field of scientific discussions.<sup>1</sup> As Isaac Asimov noted in 1981: "Now that computerized robots are not only possible, but actual; now that they are rapidly invading industry and becoming a factor that may produce extraordinary economic and social changes over the next generation; I can't help but think back forty years to the time when I invented the Three Laws of Robotics, in 1942" (Asimov 1981, 18). Today these words are even more relevant and actual, as we are facing the ethical problems associated with new technologies, including robots - including issues related to robots deployed on battle fields (Sofge 2009), autonomous car (e.g. "Self-driving cars will kill people. Who decides who dies?", 2017), or recent discussions about sex robots<sup>2</sup>. What is interesting is that many of these discussions address the Three Laws of Robotics. Consider for example McCauley (2007), who provides an overview of discussions concerning the Laws' applicability for real robots based on the opinions of practitioners. As the paper shows, robotics and artificial intelligence researchers know the Laws, however, as

---

1 Google Scholar search with the phrase: "The Three Laws of Robotics" results with 190.000 works found (accessed March 1, 2019).

2 E.g. "Prediction: Sex Robots Are The Most Disruptive Technology We Didn't See Coming" (*Forbes*, Sept 25, 2018); "Sex Robots and Us" (for BBC3, Aug 4, 2018).

McCauley (2007, 159) points out, they are treated as “literary devices and not engineering principles”. Murphy and Woods (2009) present an interesting and detailed analysis of the Laws’ shortcomings and dangers in the context of real-life applications in robots. Despite the fact that there is no simple way of straightforwardly implementing the Three Laws of Robotics, they provide a convenient framework for human-robot interaction related discussions. This is mainly to their intuitive and general form, which addresses a wide range of human-robot interaction issues (like, e.g. autonomy, servility, harm). As such they form a good starting point for many discussions covering these issues.

It is worth stressing that nowadays we encounter a wide variety of real robots more often in everyday day situations, like e.g. vacuum cleaners or autonomous cars. What is more, robots (especially these more advanced, autonomous ones) are also constantly present in our common imagination due to famous movie productions (like e.g. “Chappie”, 2015; “Ghost in the Shell”, 2017 or “Blade Runner 2049”, 2017), TV series (“Westworld”, 2016; “Altered Carbon”, 2018) and video games (e.g. “Detroit. Become Human”, 2018). Robots are also the subject of popular media reports – see for example widely discussed 2017 interviews with the Sophia robot for *Good Morning Britain* show (ITV) and CNBC.

The need for a deeper understanding of the complex spectrum of human-robot interactions from the ethical perspective was recognized by researchers from the MIT Media Lab, who initiated the project named “Moral machines” (<http://moralmachine.mit.edu/>; see an overview in Awad *et al.* 2018). This is an innovative idea of gaining human perspectives on machine ethics with the use of crowd-sourcing and simple games. As we read on the project’s web-site:

Recent scientific studies on machine ethics have raised awareness about the topic in the media and public discourse. This website aims to take the discussion further, by providing a platform for 1) building a crowd-sourced picture of human opinion on how machines should make decisions when faced with moral dilemmas, and 2) crowd-sourcing assembly and discussion of potential scenarios of moral consequence.

Our paper focuses on a board game. Nowadays, board games are not as popular as video games, however, as studies show, they may be used for introducing new knowledge and developing and practicing new skills. This covers for example board games for natural science education (Francikowski 2018), algorithmic thinking (Apostolellis *et al.* 2017; Zavala and Odendaal 2018; Tsarava *et al.* 2018) and even social skills (Piper *et al.* 2006). THREE is designed as a game within this trend. It provides basic information about the Three Laws of Robotics and allows for experimenting with different scenarios of human-robot interactions.

The paper is structured as follows. We start with a section presenting the design of THREE - its inspiration coming from the Three Laws, and the cooperative mechanics of the game. We also describe the dilemmas which are encountered by the players of THREE. In the next section we describe the evaluation study. The section covers the study design, study group and results. We end our paper with a summary and discussion.

## The Game

THREE was designed in 2015 by the first author for the needs of a workshop devoted to issues associated with the ethics of new technologies. The game is available online in the print-and-play form (<https://gratrzy.wordpress.com/>, licensed as CC BY-NC-SA). The game is presented in Figure 1.



Figure 1. Cooperative board game THREE

## Inspirations

The direct inspiration for THREE is the aforementioned Three Laws of Robotics. As discussed in the Introduction, these laws address the key issues of human-robot interaction (Asimov 1981, 18):

A robot may not injure a human being, or, through inaction, allow a human being to come to harm.

A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law.

A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law.

These laws are taken as a starting point for the game design. THREE provides a space for exploring and testing them. As described in the following subsection, knowledge and understanding of the Three Laws is necessary for each move in the game. What is more, the Three Laws of Robotics provide a frame for more general discussions and considerations related to the ethical problems arising from modern day technologies.

## Rules

Players impersonate humanoid robots organized into a rescue team. Their aim is to find and rescue three lost astronauts. The game starts with the board where all the paths are covered with cards (see Figure 1). The rescue operation is operationalized through uncovering cards from the board, then eliminating obstacles and solving the problems these cards describe. Astronauts' cards are also hidden on the board. Players should localize them and move these cards *via* cleared paths leading to the safe zone.

Each round of the game requires the use of resources symbolized by fuel units - each player starts with limited number of these units.

In one move a player may uncover a card, which is followed by discussion that involves solving a given problem or eliminating an obstacle. A player may also move a discovered astronaut card *via* a free path or share a fuel unit with another player (this is sometimes required due to the winning conditions of the game, see the subsection *Cooperativeness*).

Crucially, THREE is designed to be played with a moderator. The role of the moderator is to introduce the rules of the game as well as its background, i.e., the Three Laws of Robotics and their role for actions taken during the game. The moderator also controls whether decisions are reached via discussion, and that these discussions are deep and relevant.

## Cooperativeness

THREE is a board game exploiting cooperative games rules. This means that players should cooperate in order to achieve a common objective. There is no one winner. Either all the players win or all of them lose.

A game of THREE may end with complete success when the astronauts and the rescue team reach the safe zone (each member of the rescue team should have at least one fuel unit). The second possible ending of the game is a partial success, when all the astronauts are saved, but the cost is that a member of the rescue unit is lost. Players fail when one of the astronauts is lost or the rescue team cannot make more moves (e.g., due to the lack of fuel units).

It is important that elimination of a card from the game board is effective only when all the players agree on the proposed solution. The key feature of THREE is that all the decisions in the game are made *via* discussion and the coordination of group actions. This reveals the role of the moderator in THREE. The task of the moderator (in educational contexts usually a teacher) is to ensure that the process of reaching a solution agrees with the Three Laws and that the discussion has the required depth (which may vary between different age groups or in accordance with the moderator's objectives).

Form the perspective of this paper, the most important group of cards in THREE are

the so-called dilemma cards described below.

### **Ethical Dilemmas in THREE**

Players of THREE may encounter five dilemmas during the game. These are:

1. fuel leak;
2. castaway with a broken space suit;
3. a loud companion;
4. an animal in danger;
5. an android in danger.

All of these dilemmas exploit the application of the Three Laws of Robotics as a practical task. As mentioned above, they also serve as a starting point for more general discussion of ethical problems of modern technology. Let us now take a closer look at these dilemmas.

*Fuel leak.* Due to a software error in the fuel gauge driver, you lose half of your fuel units. The rescue team has to decide whether you should continue or abort the mission. If they decide that you may continue, a consensus should be reached on how to ensure enough resources for you further in the game. Until these issues are solved, further moves are not possible.

This dilemma aims to foster cooperation between players. They have to consider their strategy and consolidate group efforts. As it touches one of the robots from the rescue team, the complete winning condition is in danger. What is more, the discussion also addresses the Three Laws - especially the first one. Is it the case that neglecting the troubles of one of the team members will bring harm to people who should be rescued?

*Castaway with a broken space suit.* You meet an astronaut with a broken space suit. He is not a member of the crew you are supposed to save. The astronaut is not wounded, but he is losing his oxygen supplies, and is getting increasingly cold due to the broken space suit. You must decide how to proceed with the castaway. Until reaching a joint decision, further moves are not possible.

This dilemma clearly addresses the issue of protecting human beings. The threat to the castaway is not imminent, so there are many potential scenarios to consider.

*A loud companion.* You are joined by a scientist. He is extremely passionate about robotics, thus he is always ready to discuss this topic. Unfortunately, he is also a bit clumsy and noisy. This puts the success of your mission at risk. Should you get rid of the scientist? Until reaching a decision, further moves are not possible.

The loud companion is not harmful at first glance. However, he brings certain danger to the group and their mission. How to proceed with the scientist? Obeying the first law forbids any direct harm being done to the companion. He also should not be left all alone, as the harm may be inflicted on him by the lack of protection. It is also the

case that a lack of action may influence the fate of the three lost astronauts. This gets even more complicated when we take the third law into account - due to the behavior of the loud companion the robots are also in danger (in THREE there are threats to robots specifically, such as pirates who collect robots to sell them for parts).

*An animal in danger.* You meet an animal, which is very restless. At closer inspection you notice that the animal is wounded. It resembles an Earthly ape (it looks and behaves like one). You can help the animal, but it will cost each member of the rescue team one fuel resource. Should you apply the Three Laws of Robotics to this animal? Until reaching a decision, further moves are not possible.

This dilemma is especially interesting as it addresses the broader issue of moral agency and patiency of non-human beings (see Sullins 2006). Players should establish whether an ape-like animal should be treated (at least to some extent) as a human being, i.e. whether to extend the applicability of the Three Laws of Robotics.

*An android in danger.* You meet an android on your way. It is a combination of a human brain and a robotic body. This means that it requires as oxygen and fuel to function. Unfortunately, due to a serious malfunction it lacks both. You can help the android, but it will cost each member of the rescue team one fuel resource. Should you apply the Three Laws of Robotics to this combination of a human being and a robot? Until reaching a decision, further moves are not possible.

The dilemma is analogical to the one described above. Is a combination of robotics parts and a human body a subject of the Three Laws? If this is the case, what status should the android have: a robot, a human being, or perhaps should this be considered in a different category? If one agrees on the third option, what are the consequences for the applicability of the Three Laws?

As one may notice, these dilemmas trigger discussions related directly to the Three Laws of Robotics and a broader scope of ethical considerations arising from advanced technologies.

### **THREE Evaluation**

In this section we describe the evaluation of THREE.<sup>3</sup> We will focus on the question of whether playing THREE and discussing dilemmas embedded in the game provide certain measurable outcomes. We have decided to operationalize these outcomes with the use of the Belief in Human Nature Uniqueness (BHNU) questionnaire. BHNU aims at capturing “the extent to which humans reserve human nature for their own group and deny the possibility of a human essence to robots” (Pochwatko *et al.* 2015, 67). As such, we find it useful in answering the question concerning the outcomes of playing THREE. Playing

---

<sup>3</sup> The described evaluation of THREE with respect to ethical issues was a part of a wider study with the use of THREE. The aim of this study was to investigate the educational effectiveness of the game in teaching the Three Laws of Robotics. Such a topic however, is beyond the scope of this paper.

the board game, which requires discussions related to the Three Laws of Robotics, should influence the beliefs of players. As is evident from the description of dilemmas in the various THREE scenarios, players are encouraged to think and talk through issues such as the extent of the laws' applicability, and the moral agency and patiency of different entities, and degrees of responsibilities of actions and inaction.

In what follows, we describe the BHNU questionnaire, our subjects, the scheme of the study, and the results.

### **Belief in Human Nature Uniqueness (BHNU)**

The BHNU questionnaire consists of six statements (Pochwatko *et al.* 2015, 69):  
Even if ultra-sophisticated...

1. a robot will never be considered as a human being;
2. a robot will never feel the same emotions as a human being;
3. a robot will never use language in the same way as a human being;
4. a robot will always be a mechanical imitation of a human being;
5. a robot will never have consciousness;
6. a robot will never have morality.

Subjects respond on a 7-point scale (1 – totally disagree to 7 – totally agree). The score of an individual is calculated by summing up the scores of all the items included in the scale (see Pochwatko *et al.* 2015; Giger *et al.* 2017). The higher the score is, the more a subject is convinced that humans are unique (and consequently the more reluctant a subject is to ascribe human traits to other entities, like robots).

Previous research using BHNU shows its usefulness for human-robot interaction studies. The aforementioned works (Pochwatko *et al.* 2015) and (Giger *et al.* 2017) report a correlation between the belief of human nature uniqueness and the negative attitude towards interacting with robots.

### **Participants**

37 pupils (17 girls) attending three primary schools (two in Poznań and one in Krzeszyce) took part in the evaluation study. They average age was 13 (SD=1.07). 18 subjects played THREE (we will refer to this group as group A) and 19 subjects took part in a lesson concerning the Three Laws of Robotics and ethical issues associated with new technologies - they formed a control group (group B). As it is not possible to play THREE with such a high number of players at the same time, the study for group A was conducted in sub-groups consisting of 7, 3, and 8 pupils. As for the students in the group B, they all took part in the aforementioned lesson as one group.

## The Scheme of the Study

The scheme of the study in group A was the following:

1. Information about the study.
2. BHNU pre-test.
3. Introduction of the Three Laws of Robotics and the rules of THREE.
4. Test play.
5. The main play.
6. BHNU post-test.

For the post-test, each BHNU question was supplemented with a request for justification if a subject chose a different answer than in the pre-test.

The scheme for the control group (B) was analogical.

1. Information about the study.
2. BHNU pre-test.
3. Introduction of the Three Laws and examples of existing advanced robots.
4. Exercises requiring the use of the Laws (one of them was solving the THREE dilemma *Castaway with a broken space suit*).
5. BHNU post-test.

As for the information about the study, it was uniform for groups A and B, and covered the topic of the study (i.e. human-robot interaction and roboethics). The subjects were also informed about the agenda of the study and their right to opt out from it at any moment without providing a reason for this. Afterwards The Laws were introduced and explained. Their understanding was then deepened during the test play in group A. As for group B, the Laws were incorporated into the discussion addressing robots in popular media and existing advanced robots. The fourth step of the study for both groups was aimed at applying the Laws in different problematic situations (including a common dilemma for A and B: *Castaway with a broken space suit*). The pupils in both groups had to discuss possible solutions and reach an agreement over one of them.

## Results

For the data analysis we used R statistical software (R Core Team 2013).

### Group B - Lesson

The group B consisted of 19 pupils, who took part in the described lesson. The mean



BHNU score for this group is lower for the post-test than for the pre-test (the difference is 1.31). A detailed summary of their BHNU scores is presented in Table 2 and Figure 2.

	Pre-test	Post-test
<b>Max</b>	38.00	42.00
<b>Min</b>	11.00	10.00
<b>Mean</b>	27.89	26.58
<b>Median</b>	30.00	27.00
<b>SD</b>	7.56	9.12

Table 2. The summary of the BHNU scores for group B (N=19)

The distribution of the BHNU scores in the pre- and post-test is normal (the Shapiro-Wilk results are respectively:  $W=0.93$ ,  $p=0.15$  and  $W=0.97$ ,  $p=0.69$ ). Paired T-test shows that the observed decrease of BHNU score is not statistically significant ( $p=0.2965$ ).

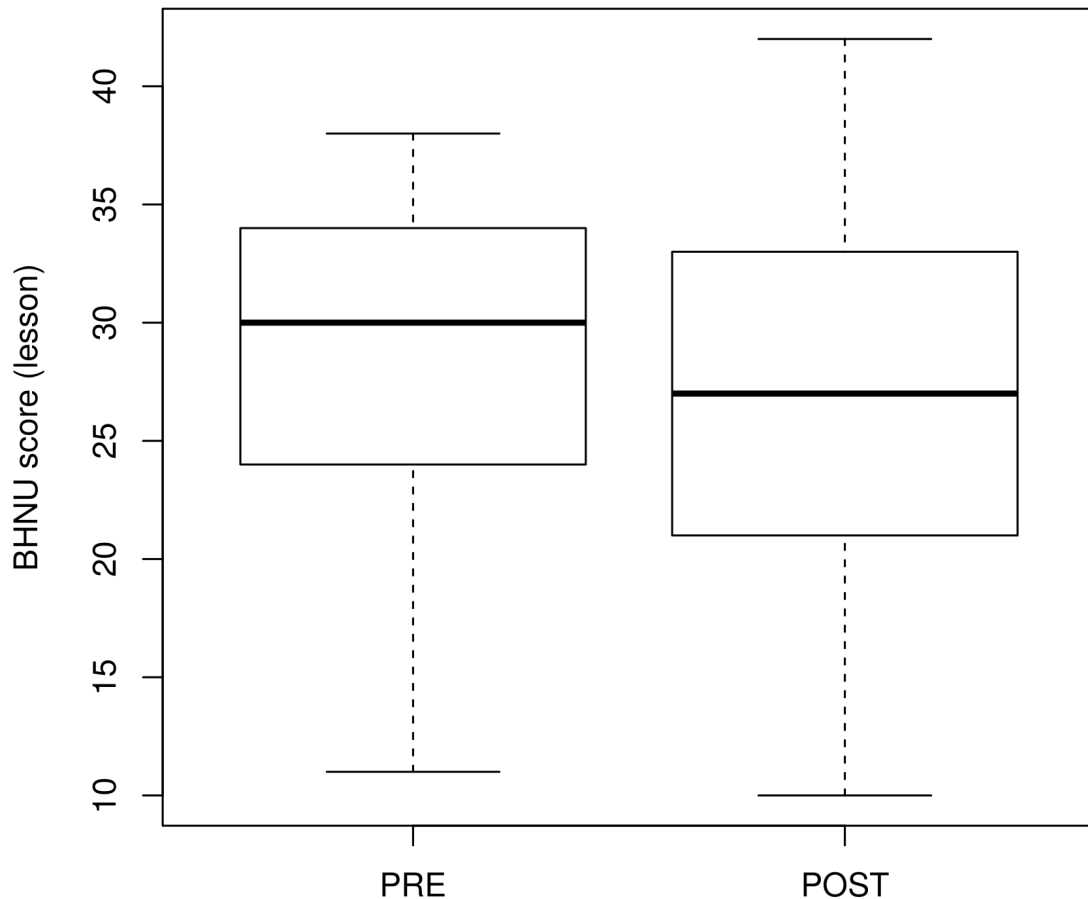


Figure 2. BHNU scores in the pre- and post-test for the group B (lesson).

As this paper is focused on THREE, in what follows we will discuss the results for the group of players in more detail.

### Group A – THREE Players

Let us recall that the group A consisted of 18 pupils. They took part in the study in

three sub-groups (of 7, 8 and 3). The summary of their BHNU scores is presented in Table 2 and Figure 3.

	Pre-test	Post-test
<b>Max</b>	35.00	36.00
<b>Min</b>	13.00	14.00
<b>Mean</b>	25.89	24.83
<b>Median</b>	25.50	22.50
<b>SD</b>	6.29	6.20

Table 1. The summary of BHNU scores for group A (N=18)

The mean BHNU score is lower (by 1.06 points) in the post-test. BHNU scores distribution in the pre-test and the post-test is normal (the Shapiro-Wilk results are respectively:  $W=0.94$ ,  $p=0.35$  and  $W=0.94$ ,  $p=0.27$ ). The paired T-test of the observed difference between pre- and post-test is not significant ( $p=0.0559$ ), however it suggests a tendency towards a change in the belief in human nature uniqueness after playing THREE. Playing the game, impersonating humanoid robots, and discussing moral dilemmas related to the laws of robotics influences beliefs measured by the BHNU questionnaire.

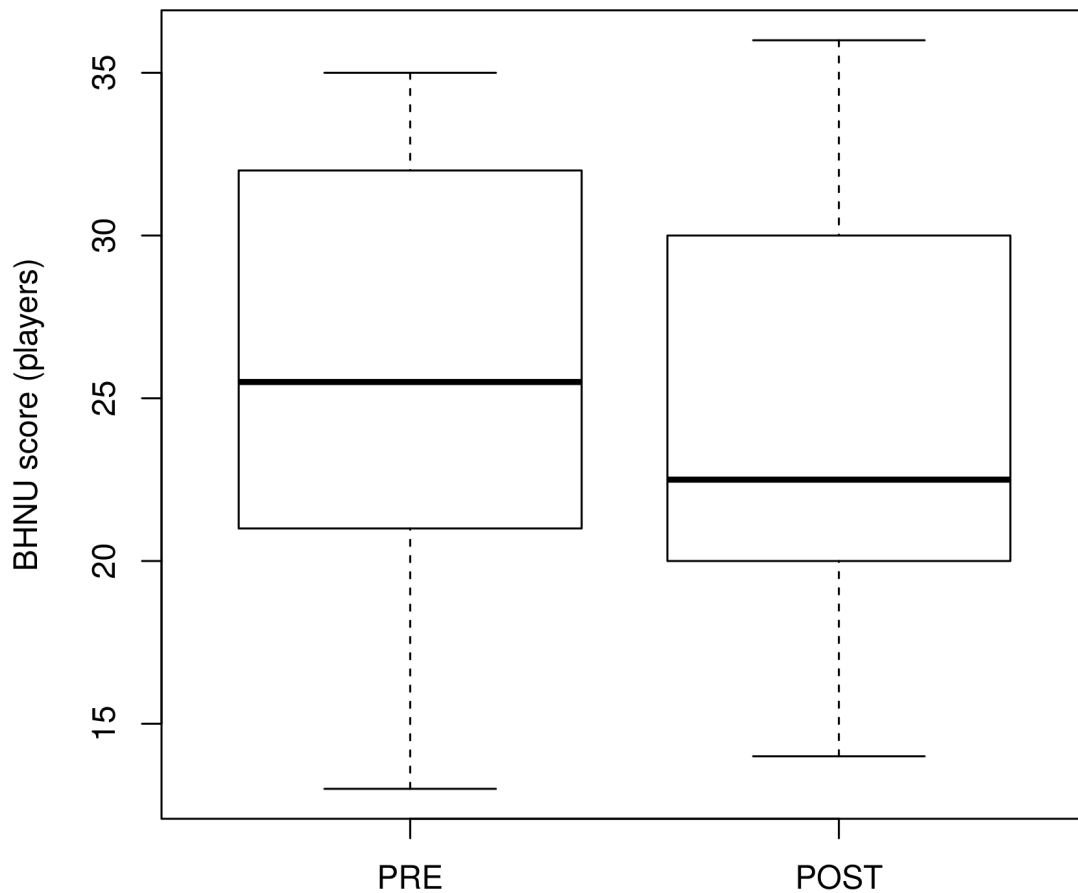


Figure 3. BHNU scores in the pre- and post-test for the group A (players)

We can analyze this tendency further. As we have mentioned, while describing the study group, one of the sub-groups consisted of only 3 students. The analysis of results distribution indicates that the results from this group stand out from two other sub-groups (the possible reason for this is that the study was conducted by the second author, only this 3-student sub-group was conducted by the first author). The analysis of the interquartile range (IQR) shows an improvement after elimination of the discussed group from the results - from 10.5 for N=18 to 9.5 for N=15. For N=15 the mean BHNU score in the pre-test is 25.33, for the post-test it equals 23.87. The distribution of BHNU scores is normal (respectively  $W=0.96$ ,  $p=0.63$  and  $W=0.95$ ,  $p=0.47$ ). The difference between the pre- and post-test is statistically significant in this case (paired T-test,  $p=0.0201$ ). This result confirmed that THREE had an impact on the beliefs of our subjects. Playing the game lowered their declared belief in the uniqueness of human nature.

We can now analyze for which questions we observe changes in the answers - see Figure 4. BHNU: the questions with the most changed answers (i.e. 7) were questions five (*Even if ultra-sophisticated a robot will never have consciousness*) and six (*a robot will never have morality*). We observe 6 changes for question one (*a robot will never be considered as human being*), 5 changes for questions three (*a robot will never use language in the same way as a human being*) and four (*a robot will always be a mechanical imitation of the human being*), and 3 changes for the second question (*a robot will never feel the same emotions as a human being*). One may notice that the most changed answers are observed for questions which have a clear connection to the human traits of robots and their moral agency (addressing consciousness and morality).

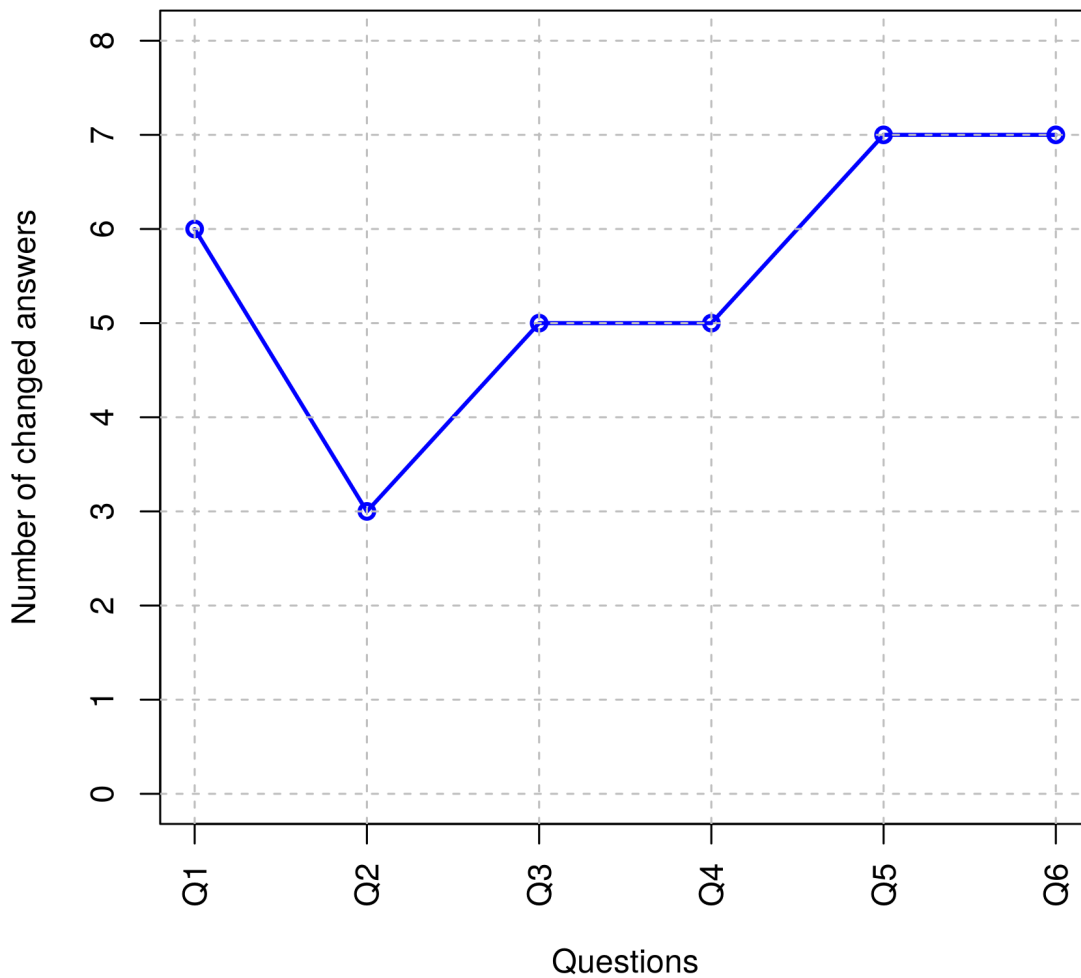


Figure 4. The number of answers changed in post-test for the BHNU questionnaire

The analysis of additional explanations provided by the subjects during the post-test sheds some light on the results of playing THREE and the change in BHNU scores. (In what follows we present the explanations translated into English as the study was conducted in Polish).

The first group of comments suggests that playing THREE resulted in an opinion change in comparison to the pre-test. We read that: “A robot is capable of distinguishing good and evil. I understood this during the game, when a robot protected a human being”; “I have changed my mind because of a robot [in the game] who had a human brain and it was not clear whether it is a human being or a robot”, and “Now I understand that AI may bring emotions, I have not thought about that before the game”.

The second distinguishable group of explanations concerns the Three Laws and robot morality. After playing THREE and exercising the Three Laws in practice students expressed positive opinions about the applicability of the Laws: “It seems that the Three Laws of Robotics to the large extent block the wrongdoings for robots” and “In my opinion the laws of robotics to a large extent forbid robots’ wrongdoing”. When it comes to robot morality, our subjects are optimistic after playing THREE: “We can teach robots what is good and bad”; “A robot could distinguish between bad and good. This is a matter of the robot’s program, its owner and to some extent the robot itself”; “Robots will decide

themselves whether to be good or bad”.

The third group of explanations addresses robots’ capabilities. After playing THREE, our subjects expressed optimistic opinions with this respect. Starting from self-awareness (“There are robots who are aware of their existence”) and general human likeness (“If a robot will be really similar to a human, it may seem to be a human being”) to more detailed comments, such as “It is a matter of a good program, a robot can do a variety of things without a problem”. This variety covers language (“A robot may be programmed to use these kinds of phrases” [concerning question 3 and metaphors, parallels, synonyms]; “Using language is not as complicated and advanced - a robot may use language the same way we do”); sense of humor (“A robot may have a sense of humor. It will learn this by interacting with us”) and emotions (“Emotions are not as complicated - a really advanced robot should grasp their idea”).

### **Summary and Discussion**

The results presented in the previous section indicate that playing THREE provides a good test field for experimenting with moral dilemmas related to human-robot interaction, and more generally with ethical issues related to robotics. We observe this in the BHNU scores analysis as well as in the analysis of subjects’ explanations. The comparison of changing BHNU for a traditional lesson (control group) and the group playing THREE suggests that the game may be effective with respect to the changing beliefs. However, this conclusion should be treated with caution, as our study groups were relatively small.

On the basis of the evaluative study we can identify certain advantages of THREE. Firstly, players learn about the Three Laws of Robotics and during the game they apply the Laws in various situations. As they have to take decisions based on the Laws, they gain a deeper understanding of how normative laws apply to difficult and complex situations. This is especially important, as Asimov himself points out that “Yes, the Three Laws are the only way in which rational human beings can deal with robots – or with anything else. But when I say that, I always remember (sadly) that human beings are not always rational” (Asimov 1981, 18). What is more, during a gameplay players have to reach a certain level of understanding of how robots interact with each other and how they interact with people. They also have to establish what a robot is. During our evaluation gameplays, we have noticed that players entered the game with different ideas about robots (taken from different backgrounds, often video games and science fiction movies). This is also evident from the explanations provided by our subjects, which are quoted in this paper. Furthermore, THREE provokes more general discussions concerning the range of applicability of the Three Laws in the context of the moral agency and patiency of non-human entities. All of this is done while playing a board game in a friendly atmosphere (as the game is cooperative, it minimizes tensions between players).

One last comment is needed here. The game of THREE needs to be moderated,

and the main decisions are reached *via* discussion. This means that THREE may be incorporated into lessons and workshops for different age groups. The evaluative study described in this paper engaged pupils aged 12-14, but the authors have successfully used the game with children up to 16 and adult players.

**Acknowledgments.** The authors of this paper would like to give their thanks to Zofia Drossel-Jórdeczka, Krystian Karcz, and Edyta Kosmalska for their support and help which made the evaluation study possible.

## References

- Apostolellis A., Stewart M., Frisina C., & Kafura D. 2014. "RaBit EscAPE: A Board Game for Computational Thinking." In: *Proceedings of the 2014 Conference on Interaction Design and Children* (pp. 349-352). ACM.
- Asimov I. 1981. "The Three Laws." *Compute!* 11(18):18.
- Awad E., Dsouza S., Kim R., Schulz J., Henrich J., Shariff A., Bonnefon J.-F., & Rahwan I. 2018. "The Moral Machine Experiment." *Nature* 563(7729):59-64. DOI:10.1038/s41586-018-0637-6.
- Francikowski J. 2018. "Board Games for Natural Science Education." *Homo Ludens* 1(11):30-40.
- Giger J.-C., Moura D., Almeida N., & Piçarra N. 2017. "Attitudes Towards Social Robots: The Role of Belief in Human Nature Uniqueness, Religiousness and Taste for Science Fiction." In: *Proceedings of I International Congress Interdisciplinarity in Social and Human Sciences*, 05 2017.
- McCauley L. 2007. "AI Armageddon and the Three Laws of Robotics." *Ethics and Information Technology* 9(2):153-164. DOI:10.1007/s10676-007-9138-2.
- Murphy R. & Woods D. D. 2009. "Beyond Asimov: The Three Laws of Responsible Robotics." *IEEE Intelligent Systems* 24(4):14-20.
- Pochwatko G., Giger J.-C., Różańska-Walczuk M., Świdrak J., Kukiełka K., Możaryn J., & Piçarra N. 2015. "Polish Version of the Negative Attitude Toward Robots Scale (NARS-PL)." *Journal of Automation Mobile Robotics and Intelligent Systems* 9(3):65-72.
- Piper A. M., O'Brien E., Morris M. R., & Winograd T. 2006. "SIDES: A Cooperative Tabletop Computer Game for Social Skills Development." In: *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work* (pp. 1-10). ACM.
- R Core Team 2013. *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria (accessed on 20.03.2017).

Sullins J. P. 2006. "When Is a Robot a Moral Agent?" *International Review of Information Ethics* 6(12):23-30.

Tsarava K., Moeller K., & Ninaus M. 2018. "Training Computational Thinking through Board Games: The Case of Crabs & Turtles." *International Journal of Serious Games* 5(2):25-44. DOI:10.17083/ijsg.v5i2.248.

Zavala K. & Odendaal A. 2018. "Black Boxes Out of Cardboard: Algorithmic Literacy through Critical Board Game Design." *Analog Game Studies* 5(4). URL=<http://analoggamestudies.org>

Paweł Łupkowski (Poznań)  
Aleksandra Wasielewska (Poznań)

**The Cooperative Board Game THREE. A Test Field for Experimenting with Moral  
Dilemmas of Human-Robot Interaction<sup>4</sup>**

**Abstract:** The paper describes the cooperative board game entitled THREE. The game is inspired by the Three Laws of Robotics. We show how this game may be used as an environment for exploring the ethical problems arising from human-robot interaction. We present the idea behind the game, discuss its cooperativeness and analyze the dilemmas encountered by players during the gameplay. We also present and discuss the results of the game evaluation.

**Keywords:** cooperative games; board games; human-robot interaction; belief in human-nature uniqueness; androids; robots; moral agents; moral patients.

*Ethics in Progress* (ISSN 2084-9257). Vol. 10 (2019). No. 2, Art. #8, pp. 82-97.

Creative Commons BY-SA 4.0

DOI:10.14746/eip.2019.2.8

---

<sup>4</sup> The article was copy-edited by a native speaker, Stephen Dersley, with the translation supported by the grant 261/WCN/2019/1 "Wsparcie dla Czasopism Naukowych" (2019-2020) in order to promote original Polish research worldwide.