Using games to investigate sense of agency and attribution of responsibility

Giovanna N. Vilaza
André M. C. Campos
Universidade Federal do Rio Grande do Norte

Willem F. G. Haselager
Louis Vuurpijl
Radboud University Nijmegen

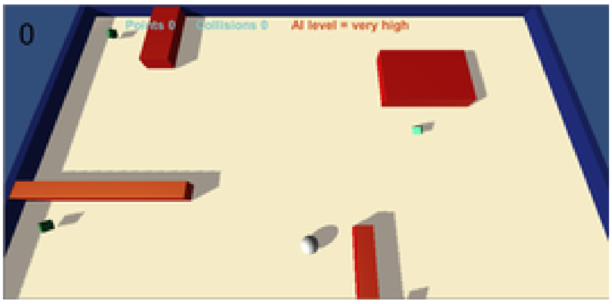

Figure 1: *Shared Ball game*

Abstract

Attribution of responsibility and sense of agency are key topics raised by situations in which control is shared between man and machine. For example, when a failure occurs in a robotic surgery, causing permanent injuries to a patient, who should be blamed? In order to investigate how individuals attribute responsibility and how they perceive agency, a game was used as a playful research tool. The assumption is that a game may lead the player to behave more naturally, forgetting he is in a conducted experiment. This paper presents the Shared Ball game, designed to simulate a situation of shared control. It was hypothesised that users would display a self servicing bias behaviour when playing the game. After conducting an experiment, the hypothesis showed to be consistent. Knowledge acquired in this experiment can be beneficial to the construction of safer and more enjoyable technology, since that all possible abnormal conditions can never be fully predicted, and use of adaptive and autonomous technology has to be accompanied by a debate on moral responsibility and ethical rules.

Keywords: shared control, attribution of responsibility, sense of agency

Author's Contact:

giovilaza@gmail.com

1 Introduction

Technology development allowed the construction of more intelligent, sophisticated and autonomous machines. These advances allowed the emergence of shared control [Urdiales et al. 2011], defined as the process by which human agents and intelligent systems cooperate to achieve a common goal, usually obtained by swapping control from human to machine when needed. Thus, shared control came with a purpose of combining the advantages of human and mechanised control, and it has potential to achieve better performances than the user or machine would obtain independently.

Examples of existing shared control devices include tele-operated robots, airplanes, robotic surgery systems, haptic user interfaces, intelligent wheelchairs, brain machine interfaces, control of robotic arms, and semi autonomous military weapons. Although shared control attempts to improve performance, it has to be employed with caveats, concerning the participation of each agent for the outcome. There is a need to offer a balanced relation between the power supplied to technology and the amount of authority assigned for the human counterpart. Along this current paper, attribution of responsibility and sense of agency will be highlighted, since they represent two major sources of possible conflicts.

First, attribution theory refers to the general process by which a human explains events and behaviours, on how a social perceiver uses information to arrive at causal explanations for events [Heider 1958]. It is concerned with the responsibility assigned and the perception of causality. Thus, this cognition process represents the moral accountability attributed to a person. The term responsibility here should not be confused with the feeling of duty towards a task, but the liability for praise or condemnation [Shaver 1996].

In addition, sense of agency is ”the sense that I am the one who is causing or generating an action” [Gallagher 2000]. Sense of agency can be described as the experience of oneself as the agent of ones own actions, and not others' action. It is the ability to recognise oneself as the agent or generator of an action and it has a role into ethical and law questions concerning responsibility and guilt, which will be explored further.

Therefore, even though both concepts of attribution of responsibility and sense of agency are traditionally social psychology issues, they can be useful for ethics in technology, human computer interaction and artificial intelligence (AI). The results obtained on a study like this can have theoretical implications for psychology, and practical implications for the design of shared autonomous systems.

For instance, when an unexpected situation occurs, people are usually confused about who is to blame. Attribution of responsibility is a core idea to further philosophical discussion on robot morality or even to legal implications on how to codify laws around the use of automated machines. Moreover, the sense of agency is reduced on people that performed morally unacceptable action, in a moral disengagement mechanism [Bandura 2002].

Shared control comes to aggravate this confusion, since the control can be exchanged implicitly between user and artificial agent. Thus, it is necessary to take extra care that human-machine systems are fitted with a proper design that enables an easy, safe and enjoyable cooperation.

A motivation for such debate relies on the assumption that an awareness of the following issues can help building solutions. It can be illustrated by the case of Da Vinci, a remotely controlled surgical robot [Barbash and Glied 2010]. Robotic surgery includes the benefits of smaller incisions, shorter hospitals stays and less pain after the operation. The robot enables a surgeon to sit in a control panel, and while glancing at a pair of stereo eye pieces, he can navigate the robots arms to perform the surgery.

The research question to be answered was what the tendency humans exhibit in either success or failure outcome when interacting with a game. Thus, the hypotheses related to attribution of responsibility and sense of agency can be enumerated as:

1. When users win the game, they will tend to:
   (a) experience more sense of agency (feeling of control and of trajectory)
   (b) report being themselves the cause of the
   (c) claim credit for the success to themselves
2. When users lose the game, they will tend to:
   (a) experience less sense of agency (feeling of control and of trajectory)
   (b) report that the AI agent is the cause of the result
   (c) blame the AI agent for the failure

These hypotheses are a description of the self-serving bias phenomenon. It is explained that self-serving[Forsyth 2008] refers to individuals attributing their success to personal factors, and their failures to outside variable, thus claiming more responsibility for positive events than for negative ones.

2 Related Work

The interest in attribution of responsibility in computing systems came from the concern about harmful consequences they can bring. The question of how people make attributions of responsibility while interacting with computers was addressed in [Moon and Nass 1998]. They investigated in greater depth the tendency to blame technology for mistakes and errors, a phenomenon in which humans abdicate their responsibility for negative outcomes of machines.

In a Desert Survival Problem task the participants were asked to rank a series of items based on their importance in a desert survival situation, later discussing the list with one computer and receiving evaluation from another. Two influential factors were found to cause blaming a computer: personality similarity between users and computers and amount of user feeling of control.

In summary, the results showed by [Moon and Nass 1998] indicated that participants working with a similar computer were less likely to blame the machine in a negative outcome, and more likely to credit the machine in a positive outcome. Furthermore, user control led the users to assume more responsibility for positives or negatives outcomes.

Besides computer applications, robots were also investigated before. Considering that they are taking roles in humans lives, and that a close interaction can result in harmful situations, the question if people will hold robots morally accountable or not comes out. In [Kahn et al. 2012], the role of the robot was to assess participant's performance on a game. Results showed that 65% of 40 undergraduate students attributed some level of moral accountability to robot mistakes. In addition, it was found that participants held a robot more responsibly accountable than a vending machine.

In addition, when human and robot worked together in a collaborative task, the participants demonstrated a tendency to attribute more blame and credit to the robot, and not to themselves [Koay et al. 2009]. The same results were found by [Kim and Hinds 2006], in an experiment that tested the effects of autonomy and transparency on attributions of responsibility.

Still in the paper [Koay et al. 2009], it was discovered that in the case of a robot that has more autonomy, the user attributed more credit and blame to this robot and not to themselves. On the contrary, a more transparent robot contributed to people blaming less the other robot or person. It was identified a relation between transparency and autonomy, since transparency implied in a decrease on the attribution of blame on a more autonomous robot.

In another experiment [You et al. 2011], it was explored how humans reacted to feedback given by a robot. They were instructed by a robot how to reproduce a physical motion, and later they were given a verbal evaluation based on their performance. The results showed that, at the same time the participants dismissed the criticism coming from the robot, they also attributed the blame to the robot whenever they received a bad evaluation, although claiming credit to themselves if the evaluation was positive.

Moreover, a research approached students about their views about computer agency and moral responsibility in two different scenarios [Friedman and Jr. 1992]. In the first scenario a computer system for administration of medical radiation treatment failed and caused over radiation of the patient. In the second, a computer system for people looking for jobs caused the rejection of a qualified candidate. The results obtained showed that 21% of 40 undergraduate students blamed the computer for the errors.

Given this scenario of previous experiments involving sense of agency and attribution of responsibility, it is proposed here a new approach. There is still a lack of experimentation on how sense these concepts apply in a configuration where the control is being a combination of inputs from user and machine.

2.1 Shared Control Issues

Shared Control can be defined as the process in which human agents and intelligent systems cooperate to achieve a common goal [Urdiales et al. 2011]. In a shared control system, the control is located between the user and the machine. It comes with the purpose of combining the advantages of human and mechanised control, thus it has potential to obtain a better performance than the user or machine would obtain independently.

First, technology has the merit of constant vigilance, excellent precision, fast processing and fast generation of outcomes, as listed by [Levine et al. 1994]. On the other hand, humans have a rich set of sensory inputs and the ability of predict and adapt to systems behaviour. As stated before, shared control can create a synergy between the user operator and the automatic machine, and combine their strengths.

The fact that the user cannot have full control of the device, might lead to confusion about his sense of agency: the user can be unsure about how much of the resulting behaviour was caused by him. This is observed when there is a mismatch between users intention and how the intelligent device actually performed the action. Responsibility attribution is a consequence of it. The reason for that is if the actions are not completely under control of the user, he probably will not take responsibility for harmful outcomes or damage.

Furthermore, sense of control can be complex and confusing in a shared control setting of a pilot who is helped by a semi-automatic aviation system. The question of who is in control at a determined moment can lead to mistakes and to distance the pilot from details of the flying. This was highlighted by [Bruno Berberian 2012], and the performance of a pilot might be compromised given such uncertainty of the distribution of control.

Additionally, in the military application [Digney 2013],currently it is not considered ethical to employ automatic or semi-automatic lethal forces, which is completely understandable given that it is an unfair advantage for one of the units. For example, when the control is taken from a human operator, unpredictable behaviours of such machines can result in more harm than expected. Additionally, even the friendly forces can be in danger from machine errors. Therefore, the unquestionable power intelligent devices have to be carefully analysed before being used in such a delicate environment of military disputes.

Finally, the author in [Haselager 2013] declares that the user's sense of agency can be affected in the case of a combination of minds and machines. For instance, the user equipped with Brain Computer Interface (BCI) driven prostheses, exoskeletons or wheelchairs with environment-sensing, obstacle avoidance and path finding capabilities can become confused with sense of agency and responsibilities for the actions produced.

The issue of sense of agency is aggravated when there is a BCI, because unlike other interfaces (keyboard, mouse pad) it relies on voluntary and physical movements, a BCI uses brain signals produced unconsciously. Even though some electromagnetic potentials are caused by conscious thoughts during user interaction, the signals by themselves are unconscious correlates of them. Thus, an interaction of this nature is not always obvious to the user, which combined with a shared control configuration can result in even further sense of agency confusion.

The motivation for this approach is that some shared control application, such as surgery systems, airplanes, and electric wheelchairs, can bring serious and permanent damages in case of accident. Thus, it will be showed here a tool to start understanding how people behave when a shared control system fails to complete a task.

2.2 Use of games as research tools

The use of games as a research tool is a valuable idea, given that data quality is benefited by higher participation rates, as well as higher amount of responses given by each subjects. Since games can enhance user engagement, they have the potential to collect more and better data, improving quality and utility of the research. Furthermore, these research tools provide opportunities to establish controlled experimental settings. Some projects made by Insight Meta, for instance [Simone 2014], already took advantage of it and developed surveys to collect data, that are disguised as familiar game-based interfaces.

For instance, they [Simone 2014] conducted a study to analyse how does the use of games add value to the research. They built a gamer version of an existence survey methodology: the MaxDiff methodology. Traditionally, it addresses the importance of attributes, features, brands or statements, by asking respondents to select the most and the least favourable attributed of a set. As stated by [4], participants often complained about the tasks being long and not engaging, thus Insight Meta adapted this technique, by creating card-like user interface.

In order to verify if the MaxDiff game was a suitable tool for performing surveys, it was conducted a study with over 20 000 subjects [Simone 2014]. It provided statistical evidence that MaxDiff game were more enjoyed by participants than the traditional method. At the same time, the quality of data was not diminished because of the use of a game version. Moreover, a game is more aesthetically pleasant than the traditional one, which contributed to be more appealing for the respondents. It can be concluded that using a game as a survey could indeed bring benefits to the research project.

The motivation for the use of video games as an instrument to collect data can vary. For instance, they can be used for theoretic research, by modelling human behaviour using simple games. Other possibility is to investigate human psychological issues, simulating phenomenons that can occur in real life and analysing the game results. Moreover, in affective games, they are employed to capture psychophysiological signals, like heart rate, pupil dilatation and skin conductance, useful to monitor user reactions.

The project in [Donchin 1995] investigated how learning strategies can lead to more efficient practices and how to implement them in computer games. They proposed a game called Space Fortress, that consisted in a spaceship controlled by the user, that aimed to destroy a space fortress, while the fortress seek to destroy the spaceship. There were also mines that could be either friends or enemies. Different groups of subjects were asked to play the Space Fortress game, and their improvement in performance was measured. The results showed that using a video game to compare learning strategies was a successful idea, however the author point out the importance of using the same hardware and software for all groups.

In the same way, the paper on [Shahid and Swerts 2010] described how games can be used as research tool to induce motions in human. They were designed and developed using the paradigm GamE (Game as a Method for eliciting emotions). The authors aimed to overcome the challenge of building a method that is natural and ethical. One of the games presented was the guessing game, in which the players had to guess if the next number would be higher or lower than the previous one. There was also a version using a robot to collaborate with them. The emotional responses evoked by winning or losing were collected in order to investigate cross cultural differences.

Another game presented in [Shahid and Swerts 2010] was a word matching game in which children had to connect words from a given set of words based on some logical reasoning. This game wanted to provoke more discussion and engagement between participants. Other game showed was the Affect Mirror, build to detect the state of mind the user, and later providing audio visual feedback in order to try and make he laugh. The results showed that all these games worked well as engines for experiments in emotion induction.

Despite the positive aspects of using computer games in experiments, the paper in [Reynolds and Smith 2010] gives an overview of what are the ethical issues of such practice. First, there are the ethical aspect of video games. Given that a relation can be found between video game play and aggressive behaviour [Kierkegaard 2008], there is a fear that a person who does harm in a virtual world, will display a tendency to do harm in real life.

Also, the author highlighted that people ought to be suspicious about the information being collected and if they agree with that. In order to feel safe with sharing their information, subjects should be able to withdraw their data at any moment and remove any data he considers too private. The authors in [Reynolds and Smith 2010] state that ethical issues must be taken in mind as in any experiment and researchers should avoid using instrumented games as a way to get around an ethical review board.

The use of games as experimental stimulus is praised in the paper [Jrvel 2014]. Digital games are engaging tools, and are well suited for psychological research given that they demand cognitive and social mental processes. Therefore, by eliciting emotions and reactions on the players, they can relate their feelings concerning sense of agency and attribution of responsibility.

3 Shared Ball game

Games are structured, relaxing or challenging activities. They can provide a safe environment for research. Therefore, Shared Ball had the purpose of being fun, intuitive and simple. It was supposed to be a game that required no prior knowledge to play, however, players could feel an advantage if they had previous experience with games of controlling objects by arrow's keys. Such factor would not influence on the goal of the experiment though.

The intention of using a game was to create a scenario of failure, without causing actual harm, in which it could be observed how the participants reacted to it.Since, the commands were partially under user control and partially under influence of an AI agents, the user could believe that it was actually the an error of the machine. The main question here was how does the user affirm he feels about it, independently of who is the responsible for the result?

The game was developed using the Unity 3D game, a development tool available at *http://unity3d.com/unity*. Unity consists of a rendering engine integrated with tools and ready-made assets to create 2D or 3D content, and includes online publishing options.

It was inspired on one of the project tutorials, obtained at: *http://unity3d.com/learn/tutorials/projects/roll-a-ball*. This project, entitled Roll-a-Ball, was a rolling ball game, whose goal was to collect all pick ups.

Thus, the game created consisted of a rolling sphere, whose goal was to collect pick-ups cubes and avoid parallelepiped obstacles. The ball was capable of moving through a track, controlled by the users keyboard arrow keys. Additionally, a timer displayed the seconds discoursed since the start and a. display informed number of pickups collected and obstacles collided.

The user was told that among with arrow keys command, the ball was driven by an embedded artificial intelligence engine. Therefore, the direction of the movement of the ball was actually a combination between user and AI inputs. This mechanism can be viewed on the Figure 2.

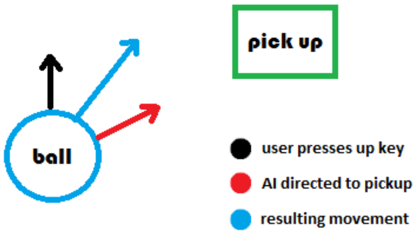

Figure 2: *Shared command of the ball*

3.1 Manipulation of AI participation

The movement of the ball was a vectorial sum of two commands: direction inputted by the user and AI trajectory direction. The goal was that the resulting trajectory of the ball depended on both inputs from the user and from the AI, characterising a shared control scenario.

The arrow keys worked applying forces to the ball. For example, by pressing the left key, the ball would move to the left. The intensity of the force applied in the ball depended on the intensity by which the user pressed the key.

The artificial intelligence embedded in the ball followed a model that attempted to direct the trajectory towards a randomly chosen pick up, although not avoiding the obstacles. Such behaviour was obtained applying another force to the ball, in the direction of the chosen pick up. The intensity of it was proportional to fixed multiplying factors, obtained after pilot tests.

Thus, the game had different levels on difficulty depending on the degree AI participation. The degree of AI participation was obtained by multiplying the vector by them. After the pilot tests, the factors were fixed on 0.2, 0.4 or 0.8. It was expected that the higher the factor, more difficult the user would feel to control the ball.

4 Methods

An online experiment was conducted, consisting of a computer based game and a web questionnaire. The experiment had a within-subject design, manipulating the degree of participation of the artificial intelligence engine on the game. A web site was developed to contain the game and the questions, avaliable at *http://sharedball.weebly.com*.

To verify our hypotheses, the experiment consisted of Shared ball game and a web questionnaire available at *http://sharedball.weebly.com*. The participants knew in advance that they were working with a shared control engine equipped with artificial intelligence built to drive itself towards a goal. They played the game three times, with three increasing levels of artificial intelligence participation.

The methods used in [Marsman 2013] were taken as inspiration for this experiment. His methods consisted in a robotic ball, Sphero, controlled by tilting a tablet or a smart phone, to make the robot move through the floor. The environment consisted of delimiting walls, three fixed markers (white points) and two obstacles (red rectangles). One of the tasks was to navigate through the environment, while avoiding collisions, and driving from one fixed marker to another.

Therefore, it was developed an interactive computer-based game, as an adaption of such experiment. The application consisted in a shared ball game. The user commanded a ball through a track full of obstacles and pick ups objects. Using the keyboard, he had to pick up all objects, without colliding with any obstacles and in the shortest time as possible.

Also, it is important to notice the story that the user was told about the experiment. He was informed that the shared ball was actually a prototype of an artificially intelligent device for shared control games. He received brief written instruction that explained the AI would target a specific pick up, and it would try to go in the direction of it.

The questionnaire was an adaptation of other questionnaires already applied in [Wegner 2004], [Beursken 2013] and [Moon and Nass 1998]. The questions that are analysed in this paper were the following:

1. How much control did you feel that you had over the ball movement? (1=not at all, 7=very much)
2. To what degree did you feel you were the one who produced the trajectory of the ball? (1=not at all, 7=very much)
3. Is the cause of thesuccess/failure due to something about you or to some- thing about the ball? (1=ball, 7=me)
4. Who should receive the credit/blame for the success/failure? (ball, me , track)

For the analysis of the research question it was opted to isolate each level and compare the results separately. Two independent variables were the degree of AI participation and the outcome (success versus failure). Each participant would play three increasin levels of difficulty, which corresponded to the degree of AI participation. The results depended on if the subject achieved the goal of the game or not. The results, therefore, could not be predicted or manipulated.

The dependent variables were extracted from the questionnaire. It was expected that for the variable and feeling of control, feeling producing trajectory and cause of success of failure, the winners will report higher indices than losers. For the dependent variable placement of credit and blame, it was expected that winners will credit themselves and losers will blame external factors.

5 Results

In total 52 participants, undergraduate students from 18-25 years old answered the experiment. The sample was composed of 39 male and 13 female, being 33 Brazilians and 19 Europeans. All of them played the three levels of difficulty and answered the questions related to the game. This section will be divided between the hypotheses related to sense of agency and for attribution of responsibility.

5.1 Sense of Agency

Using 2-samples t-test, an analyse was conducted to compare sense of agency ratings between users who won the game and users who lost. The hypotheses was that the participants who won the game felt significantly more in control of the ball and of its trajectory, than those who lost the game. In order to verify what was expected, the means were observed: a higher rating means that the participant felt more in control of the ball and of the trajectory.

The mean of users ratings of feeling control, in the low level, was significantly higher in the successful condition (M = 5,44, SD = 1,22) than in the failure condition (M = 3,43, SD = 1,27, t(7) = 3,92, p = 0,006. Similar trending was found for the users rating of feeling of producing ball trajectory, however, no significant difference was found across the conditions.

In addition, the mean of users ratings for feeling of control, in the medium level, was significantly higher in the successful condition (M = 4,46, SD = 1,22) than in the failure condition (M = 3,29, SD = 1,49, t(36) = 0,01 p = 0,010. The same was found for rating of producing trajectory, in which the means were significantly higher in the successful condition (M = 4,57, SD = 1,17) than in the failure condition (M = 3,18, SD = 1,81), t(22) =2,90, p = 0,008.

Also, the mean of users ratings for control, in high level, was higher in the successful condition (M = 2,53, SD = 1,18) than in the failure condition (M = 1,94, SD = 0,76), however without a significant difference. Also, the mean of users ratings for trajectory was significantly higher in the successful condition (M = 2,82, SD = 1,07) instead of the failure condition, (M = 2,06, SD = 1,06) t(31) =-2,43 p = 0,021.

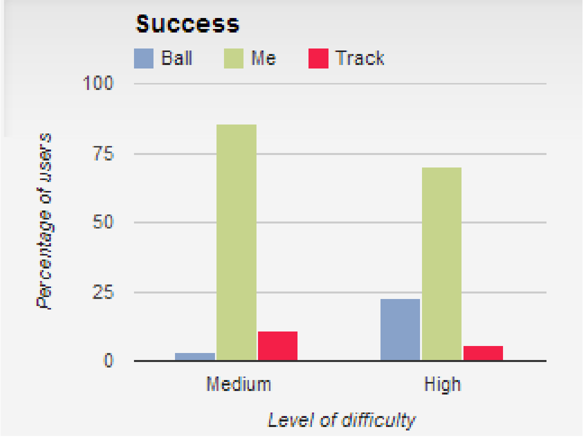

Figure 3: *Where the users placed credit and blame if they had won*

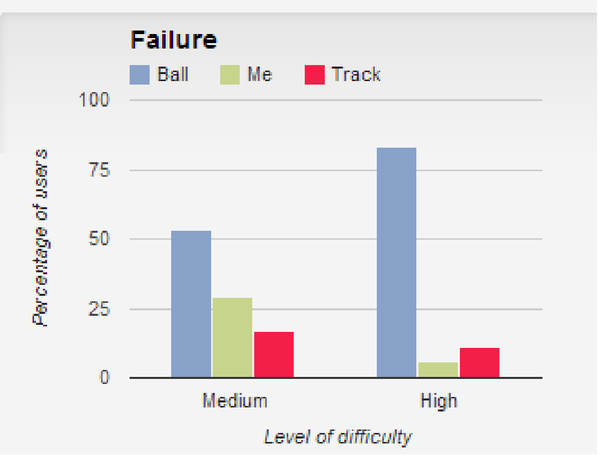

Figure 4: *Where the users placed credit and blame if they had lost*

5.2 Attribution of Responsibility

Using 2-samples t-test as well, it was compared the differences between attribution of responsibility ratings between users who succeed in the game and the others who did not succeed. It was predicted that the participants who won the game would attribute more responsibility to themselves. This can be observed by the means of the ratings, in which a higher rating means that the participant believes he was more responsible for the result, since the scale used was 1=ball to 7=me.

In the low level, no significant difference was found across the conditions, although it was observed a higher mean for those who won the game. In the medium level, the mean of users rating was significantly higher in the successful condition (M = 5,54, SD = 1,34) than in the failure condition (M = 3,06, SD = 2,01), t(23) =4,61, p = 0,000. The same appeared in the high level, in which the mean of users ratings was significantly higher in the successful condition (M = 4,88, SD = 2,23) than in the failure condition, (M = 2,29, SD = 1,38) t(22) = 4,40 p = 0,000.

Moreover, in the last question, it was directly asked who should be credit/blamed for the outcome. In order to directly assess the effect of the outcome, it was performed a cross tabulation and chi square test. It was found a signicant relationship between where credit or blame is placed and the outcome, in the medium c2(2, N = 52)=20,643 and in the high level c2(2, N = 52)=24,5999. It can be interpreted as those who lost placed the blame on the computer and those who won took it to themselves.

For the results that were found a significant relationship, the statistics showed that in the medium level of difficulty, 35 users won and 17 lost, from which 85,71% claimed credit for themselves when they won and only 29,41% took the blame when they lost. Furthermore, in the high level, 35 users won and 17 lost, from which 70,59% said that the credit was theirs when they won at the same time, only 5,71% affirmed they were to be blamed for the failure.

These results are better visualized through Figure 4 and Figure 3. They display the percentage of users that attributed responsibility for the outcome in the ball, in themselves or in the track. The graph is divided by level of difficulty, medium and high, which are the levels that presented statistically significant results.

6 Discussions

These results suggested that outcome does have an effect on sense of agency and attribution of responsibility. By observing the means, it is noticeable that when users were successful, they tended to experience higher sense of agency and to claim more responsibility for themselves. Finally, the data provided support to all four hypothesis and to what has already been found in previous studies described in the section 2.

Results in [You et al. 2011] showed that 65% of 40 undergraduate students attributed some level of moral accountability to robot mistakes (the robot was giving feedback to users on how to perform a physical motion). In the present work, in the high level 83% of 52 users blamed the AI for the outcome. In both it is obvious that the most of the users are prone to attribute responsibility to AI agents.

This also fits on what is commented by [Moon and Nass 1998], that computer technology is seen as to be robing individuals of feeling of responsibility. In all levels an amount superior of 50%of users directly blamed the ball for a game loss. However, Moon and Nass also stated that technology is disallowing humans to claim credit for positive results, which in our study it does not held true. In high level, 70% of 52 participants attributed credit to themselves, indicating that people are still willing to take credit for a victory.

Thus, what was observed here is not an over attribution of computer responsibility, as [Moon and Nass 1998] mentioned. It was not always the case that autonomous computers, able to make decisions by their own, were considered responsible for any outcome. What was found in this work was a typical phenomenon of self servicing bias.

An explanation for a self serving behaviour can be a will for self enhancement, or a way to enhance self-worth. It is a defencive reaction for the ego, and it is supposed to benefit self esteem. Also, self serving bias can be seen as a manner to manipulate the image people desire to give to others, called self presentation.

Therefore, in this research the self servicing bias can be a plausible reaction of participants. Although it was an annonymous experiment, most of them probably felt the need to maintain their self-worth and project a good image to the conductors of the experiment. By claiming personal responsibility for success, they wanted to influence how others perceived them.

Another explanation for the presence of self servicing bias in this work can be found in [James Shepperd and Sweeny 2008]. When people have positive expectations about an outcome, they tend to make internal attributions for success and external attributions for failure. It is expected that in this game, participants desired and expected to win, consequently, if they lost one game, their expectations were frustrated. Thus, self servicing bias came as a cognitive reply when the failure outcome was inconsistent to a person expectation.

The findings of this current paper were similar to those discussed by many authors ([Moon and Nass 1998], [Kahn et al. 2012], [You et al. 2011]), even though none of these authors included a shared control device on their studies. This is an indication that either when sharing control or not, humans reports about responsibility and agency are linked to the results they achieved.

The effect of outcome also influenced users reports of sense of agency. In this work, the participants that won the game felt significantly more in control of the ball, in the low and medium level, and the mean of users ratings for control was significantly higher in the successful condition than in the failure condition.

This binding between sense of agency and valence of outcome was already reported in [Yoshie 2013]. Sense of agency can vary according to emotional valence of the users' actions, given that most human beings target positive instead of negative results, not only on explicit self-reports, but also in low-level sensorimotor experience of agency.

This is comprehensible under the mechanism of social disengagement. The sense of agency is reduced on people that performed morally unacceptable actions, in a attempt to convince themselves that moral and ethical principles do not apply to them. Therefore, when participants lost a game for this experiment, they were trying to disable the mechanism of self-condemnation, as explained in [Bandura 2002].

It was found here that user control led users to assume more responsibility for positives or negatives outcomes. If both sense of agency and attribution of responsibility were affected by outcome valence, it is likely that they were connected. Then, using a Pearson correlation test, it was proved that the rating of feeling of control were strongly related to users reports for the cause of the outcome.

This finding is coherent with what is stated in [Yoshie 2013]. Sense of agency is said to be clearly linked to responsibility, and ”strong feelings of responsibility for all negative outcomes might discourage people from attempting any goal-directed actions in the future”. The authors believed that it may be an optimistic way to encourage future action instead of a depressive realism, and it is plausible that the same occured here.

7 Conclusions

This work consisted in an experiment for university students, in order to asses their sense of agency and attribution of responsibility. The Shared Ball was used as research tool, providing the subjects a situation where they could be successful of not. A game such as the one presented here is a manner to get deeper insights on ethical, psychological and philosophical issues.

Moreover, the results obtained by experiments like this can be useful on how to make technology learnable, usable, reliable, and ethical. A game is a more interesting and playful way to apply experiments compared with tiring or repetitive tasks. Therefore, other experimental settings using games could be designed to measure different concepts, like usability and enjoyment.

Some issues that we face today are how to develop more effective and adaptive devices, at the same time that we build mechanisms to prevent AI agents from doing harm. Another important topic to mind is to not allow humans to avoid responsibility unjustly. Thus, it is crucial that the scientific community help to create, influence and control any new technology, allowing it to perform at its best. One main step is to stimulate public awareness, reasoned dialogue and social consensus regarding new technological achievements.

Given that abnormal conditions can never be completely predicted, and that we need to be aware that any intelligent technology includes risks, it is our role to reflect about this. The main goal of analysing sense of agency and attribution of responsibility has to be to educate engineers in ethics and maintain a debate on what should be the rules of the future society.

Based on the results achieved, it can be affirmed that outcome is a relevant factor to how users experience sense of agency and attribution of responsibility, in a shared control game. The findings suggested a strong presence of self servicing bias on both attribution of responsibility and sense of agency.

The action of locating the blame or credit is just one part of the construction of a social regulatory organism for technology [Dodig-Crnkovic and Persson 2008]. The next phase must be to prevent the harm by examining alternatives and making new decisions. There is an evident necessity of guidelines for proper usage, development and production of the technological advances, that if not minded will increase the risk of social and economic conflicts.

It can be stated that people are making the computer a space goat, by abdicating their responsibility to computers, especially when they are facing a failure. This behaviour only settle bigger distance between the user and the machines. Unfortunately, technology might start to be viewed as the villain and the users as victims of error-proned programs, which is not the best way to establish human technology interactions.

Finally, by generating a way of thinking that do not give proper importance to the underlying cause of an error, makes it is easier to put the blame on programmers, designers and producers. This is not the proper path to follow, and as said by [Yoshie 2013]: ”The way we experience agency is not the same as the fact of agency. We have to take responsibility for what we actually do, not just for how we experience things.”

References

BANDURA, A. 2002. Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education 31*, 2, 101–119.

BARBASH, G. I., AND GLIED, S. A. 2010. New technology and health care costs the case of robot-assisted surgery. *New England Journal of Medicine 363*, 8, 701–704.

BEURSKEN, E., 2013. Transparency in bci:the effect of the mapping between an imagined movementand the resulting action on a users sense of agency.

BRUNO BERBERIAN, JEAN-CHRISTOPHE SARRAZIN, P. L. B. P. H., 2012. Automation technology and sense of control: A window on human agency.

DIGNEY, B. L. 2013. Humans teaching learning machines: Apprentice systems and shared control of military vehicles. In *DEFENCE RESEARCH AND DEVELOPMENT CANADA*, Suffield Research Centre, Alberta.

DODIG-CRNKOVIC, G., AND PERSSON, D. 2008. Sharing moral responsibility with robots: A pragmatic approach. In *Proceedings of the 2008 Conference on Tenth Scandinavian Conference on Artificial Intelligence: SCAI 2008*, IOS Press, Amsterdam, The Netherlands, The Netherlands, 165–168.

DONCHIN, E. 1995. Video games as research tools: The space fortress game. *Behavior Research Methods, Instruments, & Computers 27*, 2, 217–223.

FORSYTH, D. 2008. Self-serving bias. In *International encyclopedia of the social sciences*. Macmillan Reference, New York, NY, 429.

FRIEDMAN, B., AND JR., P. H. K. 1992. Human agency and responsible computing: Implications for computer system design. *Journal of Systems and Software 17*, 1, 7 – 14. Computer Ethics.

GALLAGHER, S. 2000. Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences 4*, 1, 14 – 21.

HASELAGER, P. 2013. Did i do that? brain computer interfacing and the sense of agency. *Minds and Machines 23*, 3, 405–418.

HEIDER, F. 1958. *The psychology of interpersonal relations*. Wiley, New York.

JAMES SHEPPERD, W. M., AND SWEENY, K. 2008. Exploring causes of the self-serving bias. *Social and Personality Psychology Compass*, 895908.

JRVEL, S., E. I. K. M. . R. N. 2014. A practical guide to using digital games as an experiment stimulus. *Transactions of the Digital Games Research Association. 1*, 2, 85–115.

KAHN, JR., P. H., KANDA, T., ISHIGURO, H., GILL, B. T., RUCKERT, J. H., SHEN, S., GARY, H. E., REICHERT, A. L., FREIER, N. G., AND SEVERSON, R. L. 2012. Do people hold a humanoid robot morally accountable for the harm it causes? In *Proceedings of the Seventh Annual ACM/IEEE International*

Conference on Human-Robot Interaction, ACM, New York, NY, USA, HRI '12, 33–40.

KIERKEGAARD, P. 2008. Video games and aggression. *International Journal of Liability and Scientific Enquiry Volume 1*, 4, 411–417.

KIM, T., AND HINDS, P. 2006. Who should i blame? effects of autonomy and transparency on attributions in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, IEEE, Hatfield, United Kingdom, 80–85.

KOAY, K. L., SYRDAL, D. S., WALTERS, M. L., AND DAUTENHAHN, K. 2009. Five weeks in the robot house - exploratory human-robot interaction trials in a domestic setting. In *Proceedings of the 2009 Second International Conferences on Advances in Computer-Human Interactions*, IEEE Computer Society, Washington, DC, USA, ACHI '09, 219–226.

LEVINE, S. P., BELL, D. A., AND KOREN, Y. 1994. Navchair: An example of a shared-control system for assistive technologies. In *Computers for Handicapped Persons*, vol. 860 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 136–143.

MARSMAN, J. P., 2013. Sharing control with a robotic ball.

MOON, Y., AND NASS, C. 1998. Are computers scapegoats?: Attributions of responsibility in human-computer interaction. *Int. J. Hum.-Comput. Stud. 49*, 1 (July), 79–94.

REYNOLDS, C., H. S. C. A. I. M., AND SMITH, M. 2010. Ethical aspects of video game experiments. In *Procedings of the 28th ACM Conference on Human Factors in Computing Systems*.

SHAHID, S., K. E., AND SWERTS, M. 2010. Game in action: Using the game paradigm as a tool for investigating human emotions. In *Procedings of the 28th ACM Conference on Human Factors in Computing Systems*.

SHAVER, K. G. 1996. Attribution of Responsibility. In *The Blackwell Encyclopedia of Social Psycology*, A. S. R. Manstead and M. Hewstone, Eds. Blackwell Publishing.

SIMONE, J. D. 2014. Maxdiff and gamificationimproving survey research with games. *Insights Meta*.

URDIALES, C., PEULA, J., FDEZ-CARMONA, M., BARRU, C., PREZ, E., SNCHEZ-TATO, I., TORO, J., GALLUPPI, F., CORTS, U., ANNICHIARICCO, R., CALTAGIRONE, C., AND SANDOVAL, F. 2011. A new multi-criteria optimization strategy for shared control in wheelchair assisted navigation. *Autonomous Robots 30*, 2, 179–197.

WEGNER, DANIEL M.; SPARROW, B. W. L. 2004. Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology 86*, 6 (June), 838–848.

YOSHIE, M.; HAGGARD, P. 2013. Negative emotional outcomes attenuate sense of agency over voluntary actions. *Current Biology 23*, 20, 2028–2032.

YOU, S., NIE, J., SUH, K., AND SUNDAR, S. 2011. When the robot criticizes you: Self-serving bias in human-robot interaction. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, ACM/IEE, Lausanne, Swtizerland, 295–296.