Master Thesis

Human Behavior towards Virtual Humans

Rens Hoegen

Human Media Interaction
&
USC Institute for Creative Technologies

Examination Committee:
Dirk Heylen
Rieks op den Akker
Merijn Bruijnes
Jonathan Gratch

08-Jun-2015

UNIVERSITY OF TWENTE.

USC Institute for Creative Technologies

Acknowledgments

I would like to thank my supervisors, Dirk Heylen, Rieks op den Akker, Merijn Bruijnes and Jonathan Gratch for providing me the opportunity to do this amazing project.

Furthermore I want to thank everyone who supported me throughout my stay in the United States and made it so much more enjoyable:

Alesia, Ali, Amir, Amanda, Anna, Antionne, Brechtje, Celso, Christopher, Christopher, Claudia, Cyrus, Danielle, Diane, Ed, Eli, Elnaz, Emmanuel, Gale, Giota, Ina, Jann, Jeroen, Jill, Johnathan, Justin, Kevin, Koki, Laura, Maarten, Marcie, Maryam, Maryam, Moitreya, Ondrej, Pete, Rachel, Ramesh, Reihane, Saghi, Sayan, Shaun, Sin-Hwa, Stefan, Steve, Tatiana, William, Zahra.

You guys are awesome!

TABLE OF CONTENTS

1 INTRODUCTION

During this project I studied how humans interact in an "Iterated Prisoner's Dilemma" with other human opponents and "virtual human" opponents. The goal of this study was to compare the behavior people display against these Virtual Humans, with the behavior shown against other people. The virtual humans that were used during these studies are three-dimensional embodied agents rendered on a computer and are capable of showing facial expressions and other animations. This study was done using an existing framework that allows participants to play the Prisoner's Dilemma against each other. One necessary aspect of the project was to add the functionality to play against Virtual Humans to framework. Two other economic negotiation games were also implemented within the framework. These games were played against a Virtual Human exclusively.

This project was done at the Institute for Creative Technologies (ICT), a research institute affiliated with the University of Southern California. Much research into integrated virtual humans has been done at ICT [1].

1.1 PROJECT GOALS

The first thing that needed to be done during the project was to improve the framework that was already used for running experiments, the "Web Gym" framework. During these experiments participants would play specific games against each other. The framework was used during studies where participants played games within a web browser and allowed the participants to communicate with each other through their webcam. Most of the games that are played within the framework are "negotiation games", such as the well-known prisoner's dilemma and the ultimatum game that is often used in economic experiments. Along with the decisions that participants made in the game, the framework also stored the videos of their webcams, the expressivity of the participant in these videos could then be analyzed using specific software.

One addition to the framework was the ability for participants to play games against a virtual human. Because the framework was originally created for studies involving games played between two human participants, changes needed to be made to the system in order to support games between a human and virtual human participants.

The system already contained an "Iterated Prisoner's Dilemma" game and a simplified poker game. The prisoner's dilemma was adapted by allowing participants to play against a virtual human instead of a human opponent. New games made specifically to be played against virtual humans were created for the framework as well. The framework was built in a generic way, so it would be relatively simple to add new functionality and allow the system to be used for future studies as well.

After adding this new functionality the improved framework was used as a platform for studies. Two studies were performed using the framework. During both studies participants played two different games using the framework. Some of these games were played against a virtual human while others against human opponents.

In the first study participants played the Iterated Prisoner's Dilemma against other human opponents or against a virtual human. The goal of this study was to determine whether there was a difference in social behavior participants showed against a virtual human compared to playing against a real human. Therefore I will refer to this study as the "VH study" in this report.

Data was gathered for several other studies during the experiment for the VH study as well. One aspect that was looked at was the concept of "emotional regulation", this refers to the processes that people might use to influence the emotions they have and how they experience and express these emotions [2]. A previous study by van Vroonhoven [3] showed that participants with a high emotional regulation score did better in the prisoner's dilemma than other participants did. In order to gain more insight into this, participant's emotional regulation was manipulated using a set of instructions as proposed by Grecucci [4].

In addition to the Prisoner's Dilemma, participants played a second scenario, the "Ultimatum" game [5]. The ultimatum game was based on a previous study [6], the game that was used in the previous study was considered too complicated however. It also simply used text and a pictogram to simulate a virtual agent. The new game created for the framework contained an actual virtual human and a simplified version of the game, so participants would have an easier time understanding it.

In the second study done for this project, participants played the iterated prisoner's dilemma again. This new run of the prisoner's dilemma was done because a preliminary analysis of expressivity of participants showed expressivity was lower against the virtual human than it was against real humans. One possible explanation for this is that participants communicate less if their opponent shows less expressivity, because the Virtual Human only displays an expression at a specific point in a round. Therefore in order to see if the expressivity of an opponent influences the participant, half of the participant pairs would only see a single still image of their opponents instead of a video during this study. If the opponent influences the participant's expressivity this would show in a lower expressivity score for these participants. I will refer to this study as the "stills study" within this report. In addition to the stills and video participant pairs, some participants were paired with a virtual human opponent, as such there were three groups in the stills study.

Just like during the VH study, participants played an additional game in the experiment for the still study as well. This was the "investment" game, which was made to be played against a Virtual Human exclusively. The investment game was based on a previous version which used a virtual agent. This game was played twice during the experiment, against two different virtual humans.

Finally the results of the studies were analyzed. The prisoner's dilemma games is the only game that allowed me to directly compare behavior of participants, as this is the only game where participants could either play against a virtual human or a real human opponent. Therefore the analysis mostly focused on this game.

1.2 Outline

This thesis is structured as follows. In chapter 2 the related work will be examined. I will discuss the previous studies using economic negotiation games, which the games within the framework were based on. I will touch on some other studies using virtual humans, both employed as agents (controlled by computers) or as avatars (controlled by humans). Based on this related work I will formulate the hypotheses of this project at the end of that chapter.

Chapter 3 will consist of a description of the framework, both the original framework will be described as well as the additions made to it during this assignment, such as the addition of a virtual human opponent and the new games, the ultimatum and investment games.

Chapter 4 describes the experimental setup and results for the VH study, chapter 5 does the same for the stills study. In chapter 6 the implications of the results will be discussed, the overall conclusion of this project will be given as well. Lastly details will be given on possible future work based on the obtained data and results of these studies.

2 RELATED WORK

The games played within the framework are all negotiation games, commonly used in game theory studies. I will also discuss previous studies with virtual agents, focusing specifically on the way these agents are presented to the participants. Lastly, based on the literature, the hypotheses of my project will be formulated at the end of this chapter.

2.1 NEGOTIATION GAMES

During this project I worked with three different negotiation games; the iterated prisoner's dilemma which was used in both studies. In addition to that I created the ultimatum game and the investment game for two other studies.

2.1.1 Iterated prisoner's dilemma

The original prisoner's dilemma [7] is a game often analyzed within the game theory field. The game is played with two players and takes place in a prison, both participants play the role of an arrested criminal. The participants are in solitary confinement and have no means of speaking with each other. Because there is not enough evidence to convict the pair, the prosecutor offers each prisoner a bargain: They are given the opportunity to either betray the other by testifying to the committed crime or staying silent. Based on the choices of both participants there are several possible outcomes: If participant A and B betray each other both of them serve 2 years in prison. If participant A betrays B but B stays silent, A will be set free while B will server 3 years in prison and vice versa. If both A and B remain silent they will both only serve 1 year in prison.

Many different variations of the game exist, although the core choices of cooperating with each other (staying silent) or defecting (betraying your opponent) should be the same. Additionally the pay-off for choices should be in the following order: The best individual score can be obtained by defecting upon a cooperating participant, the second best by both cooperating, then both defecting and being defected upon when choosing to cooperate will get the lowest score. This means that if participants would pick their choices purely out of self-interest, defecting upon the participant would be best. However if both participants do so, they would get the second worst individual outcome, thus making the cooperation option viable. Additionally participants tend to have a bias towards cooperative behavior.

For this project an iterated prisoner's dilemma was used. The main difference between the standard and iterated version of the prisoner's dilemma is that there are multiple rounds to the iterated version. Within an iterated prisoner's dilemma the choices made during a round might influence future rounds, adding another layer of strategy to the game. The tit-for-tat strategy is a strategy commonly used in the iterated game, when using this strategy participants will make the same decision as their opponent did the previous round. By doing this the player will reward their opponent for cooperating by also cooperating, while punishing their opponent if they chose to defect upon them.

2.1.2 Ultimatum

The Ultimatum is a simple game often used in economic experiments. It is played by two players, one player takes the role of the proposer and the other players takes the role of the respondent. When the game starts the proposer will receive several items and gets the opportunity to propose how to divide these items between the two participants. If the respondent accepts the players will receive the items according to the proposal, however if the respondent rejects neither player will get anything.

Although the game is commonly only played once, within the web gym framework it will be possible to go through multiple rounds of proposals. In this case the idea behind rejecting an offer will usually be to try and push the proposer to give better offers. Additional rules can also be introduced to the Ultimatum, such as giving specific items specific values that are different between both participants.

2.1.3 Investment game

The investment game built during this study was first described as the "trust game" by Berg et al. [8]. This game used a similar setup as the ultimatum. It's a game with two participants and the objective is that the participants are playing for items with a specific value to participants (such as money). During the game, one participant will take the role of investor, this player will receive a sum of money at the start of the game that he can then invest to a trustee. The trustee is the other player of the game, although this will usually be someone who is part of the experiment. Once the investor decides how much money to invest and hands this over to the trustee, the invested money will be multiplied by three. After this the trustee will return some money. This return can either be fair, where the trustee return more than what the investor originally invested, or unfair where the return is less. The return amounts are variable, but a fair return will always be more than the initial investment, while the unfair return is always less.

The investment game can be played for multiple rounds as well. In this case investors could believe that by changing their investments they could influence the behavior of the trustee.

2.2 VIRTUAL HUMAN STUDIES

Within this report the term virtual human refers to an embodied virtual agent capable of displaying facial expressions. An embodied agent can be anything from a simple series of static images of the agent, to a complete 3D model capable of expressing itself using animations. Applications of these types of embodied virtual humans date at least as far back as the late 90s [9] [10] [11] virtual humans have been the focus of studies for a long time [12]. Many of these studies focus on the way virtual humans behave by creating behavioral models for virtual humans [13], or by studying means to allow computers to understand humans better [14].

Other studies instead focus on how these behavioral models influences the people the VH is interacting with. Facial expressions and display of emotion are important aspects in these studies. For example Ruttkay et al. argue that it might be necessary to model imperfections in the behavior of virtual humans in order to keep the interaction with the VH engaging [15].

De Melo has done several studies were he looked at the effect of behavioral models of virtual humans on participants of a human-agent negotiation study. In one of these he looked at the way emotion display of an agent influenced participants in a prisoner's dilemma [16]. The result of this study showed that the emotion display of the agent influenced the user's decisions, a user would concede more to an angry agent than to a happy one. In a follow up experiment to this de Melo also manipulated the participants' beliefs on how the agent was controlled [17]. To some participants the agent was introduced as an agent using artificial intelligence, while to other participants it was introduced as an "avatar" that was controlled by a person. Two experiments where participants played a social dilemma were done, one with a Prisoner's Dilemma and one with an Ultimatum game. This study showed that the effect of the emotion display depended on the perceived agency of the agent, as the effect was stronger in the avatar agent than it was in the computer controlled one.

2.2.1 Social behavior towards virtual agents

Many theories exist that try to explain how people show social behavior towards a virtual agent. One well known theory on the displayed social behavior of people when interacting with computers is the Media Equation by Reeves and Nass [18]. They claimed that the responses of people to computers would be the same as responses to humans, if these computers use human-like social cues. The reason that this happens is because over time people developed automatic responses to social cues. This allows people to react to these cues unconsciously. Therefore a human would respond to social cues no matter if these cues are shown by either a computer or by other real humans.

Nass often advocated a strict interpretation of the media equation [19]. He argued that responses towards computers would in fact be equivalent to responses to humans, if the computer incorporated social cues. Nass named this concept "Ethopoeia". He argued that these type of computers form a different group in between anthropomorphic objects (where people truly have the belief that they are interacting with a human) and cherished objects (which receive special treatment however are not treated like humans). When humans interact with these computers they would treat them equivalently to humans, even though they realize that it is not a real human.

However if one were to replace the 'equivalent' in this version of Nass's media equation with a 'less than' a more nuanced picture is painted. In this case the social influence of a computer would not be equal to that of humans, however depending on specific aspects this influence might get closer to the social influence of a real human. For example, Blascovich [20] argues that the social influence of a virtual agent will increase based on the perceived realism and its "agency". Agency in this case refers to the perceived sentience or free will of an agent.

Blacovich's view is supported by the study of de Melo et al. [17] in which the sense agency was manipulated. This was done by comparing virtual humans that were either computer controlled agents (i.e. controlled by computers) or avatars (i.e. controlled by humans) with each other. De Melo's experiments showed that if a virtual human showed a specific facial expression participants would cooperate more with that VH. However this difference in cooperation only was significant for the avatar condition. Thus showing that the way participants view a virtual human influences their behavior.

A study where participants played against humans and avatars was done by Riedl et al. [21]. They found that participants showed similar trust behavior between humans and avatars. However through neuroimaging it was shown that despite these similar responses, human and avatar opponents were processed differently in the brain. A study by Krach et al. [22] another condition was introduced, besides computers and humans, participants played against robots. The results of this study showed that participants experienced more fun and competition during the interaction as the human-like features of their partners increased.

It has been argued by van Kleef et al. that the concept of facial expressions within economic games can serve as automatic elicitors of social behavior [23], based on this de Melo et al. proposes that VHs can exploit these cues [24]. These studies form the basis of my project. People's interactions with virtual humans will be compared to their interactions with humans, in order to gain further insight into human-virtual human interaction.

2.3 Hypotheses

This project builds on the latest findings of de Melo et al. [25]. In their study, people played an iterated prisoner's dilemma with a VH that played tit-for-tat and expressed emotion. In one condition, participants believed the agent's choices and emotions were selected by another participant. In the other, they believed they were generated by a computer "programmed to behave like a human." In either case, players could "send" emotional expressions to the other player along with their choice in the game. The tit-for-tat behavior and the pattern of emotional expressions were both chosen to maximize the amount of cooperation shown by participants. Nonetheless, participants made less cooperative choices and "sent" fewer positive and more neutral expressions when they believed they were playing a computer opponent. Based on these findings, the following hypotheses are formulated:

H1: *Participants will cooperate significantly more with other human players than VH. More specifically, people will be (H1a) more willing to try to exploit a VH, (H1b) more willing to persist in exploiting a VH, and (H1c) more willing to forgive humans following exploitation.*

H2: *Participants will show more cooperative facial expressions to human players compared with VH. Specifically, people will (H2a) show more joy to human players and (H2b) show more neutral expressions to VH.*

3 THE WEB GYM FRAMEWORK

The framework that I worked with during this project is called the “Web Gym” framework. New functionality was added to the framework during my final project and studies were done using this version of the framework. Within this chapter I describe the original framework. Following this I describe the new additions made to the system; the virtual human and the ultimatum and investment games, as shown in Figure 1.

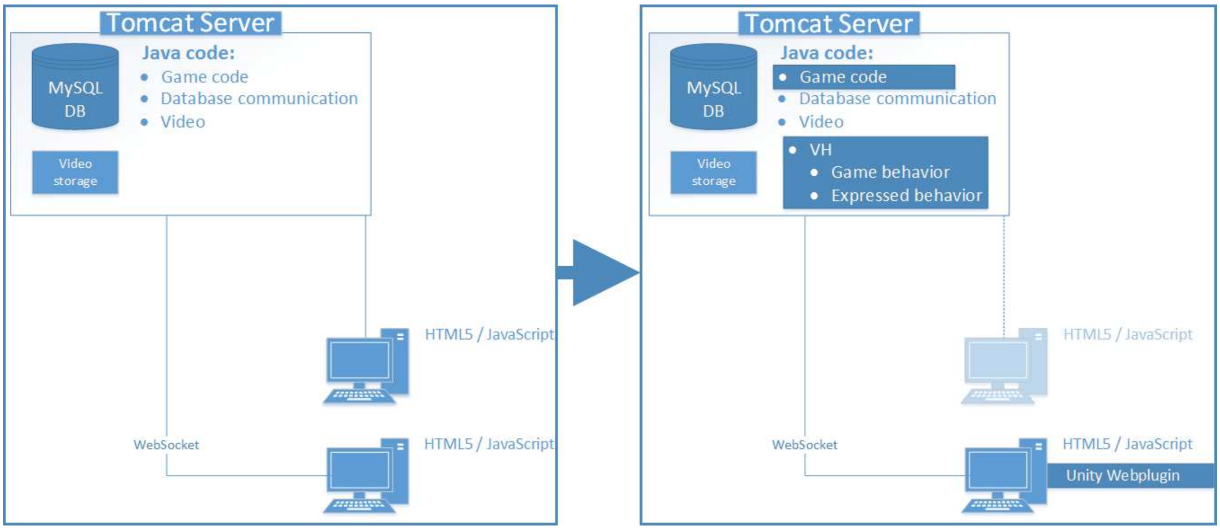

Figure 1. The model of the framework. On the left the original version, the right model shows it with the new additions: New game code, the virtual human (using the Unity plugin). Because of the addition of a VH, the second client is now optional.

3.1 THE ORIGINAL FRAMEWORK

This first version of the web gym framework was built by van Vroonhoven [3] who built this system as the final project of his master’s program. This system contained two games, a simplified poker game and an iterated prisoner’s dilemma. The poker game was built specifically for van Vroonhoven’s study and was not used during this project, while the iterated prisoner’s dilemma was used in both of the studies done with the improved system.

The system has been set up according to a client-server model, the server uses Apache Tomcat[1] a web server that uses specific java specifications, such as JSP to generate web pages and WebSocket for communication between server and clients. As this is a JSP server, most of the code running on the server is Java-based.

[1] http://tomcat.apache.org/

Participants can connect to the server as clients, through their internet browser. Once a client accesses the server by navigating to the specific URL of the server, the server will set up a WebSocket connection with the client. Through this connection the server can communicate with the client (e.g. it can send a message to the client to start a new round within a game) as well as the other way around (e.g. the client can send the choice a user made during the game to the server), most of this communication occurs while the participants are playing a game. The games that the participants play during a study use HTML5 and Javascript in order to display the game and show simple animations during game events. Most of these scripts are called when the server sends specific messages through the Websocket. This WebSocket service is set up by the Javascript code on the client side and by java code on the server side.

While doing a study, the server can direct the client to several different pages, usually these are specific games that the participants will need to play. The system can redirect participants to a Qualtrics questionnaire[2] as well, in order to gather self-reported data. When the participants start a game, they will usually first need to read some instructions, following this they can start the game. Because the system is set up for games between two human participants, the system will at this point try to match the participant up with another participant. If no other participant is found, the participant will have to wait until another participant tries to start the game. After matching up with an opponent the game will start. Each decision a participant makes will be sent to the server, which will then broadcast it to the opponent, as such all communication between the participants will have to go through the server first.

3.1.1 Iterated prisoner's dilemma

The iterated prisoner's dilemma used within the framework has been modeled after the British gameshow "Golden Balls". The framework used a prisoner's dilemma scenario, which had the participants invest in a specific project. The choices to invest in a specific project however did not have the same positive and negative connotation that Cooperate or Defect have.

Within a specific segment of the Golden Balls gameshow, two participants would be offered a large sum of money, each of the participants was then allowed to respond to this offer by either proposing to "Split" the money with the other participant or by trying to "Steal" the money from their opponent. As such the choice to Split the money corresponds to choosing to cooperate within the prisoner's dilemma and choosing to Steal the money corresponds to defecting. Thus, Split has the same positive connotation as Cooperate and Steal the same as Defecting.

The general flow of the game is as follows; first four golden balls will appear on-screen, two in the front and two in the background. The Split and Steal choices will be shown to both participants with a small animation. These choices are represented as the front two golden balls opening and display the split and steal choices to both participants, in addition to a participant's own choices the opponents chance to choose is also represented by the two smaller balls in the background opening up. The participants will then both have the opportunity to make their decision at the same time. When a participant makes a decision, their opponent will get notified of this with a small animation of their opponent's balls closing.

---

[2] http://www.qualtrics.com/

Once both participants have made a decision, the results will be displayed to both players, by another short animation. After this the next round begins.

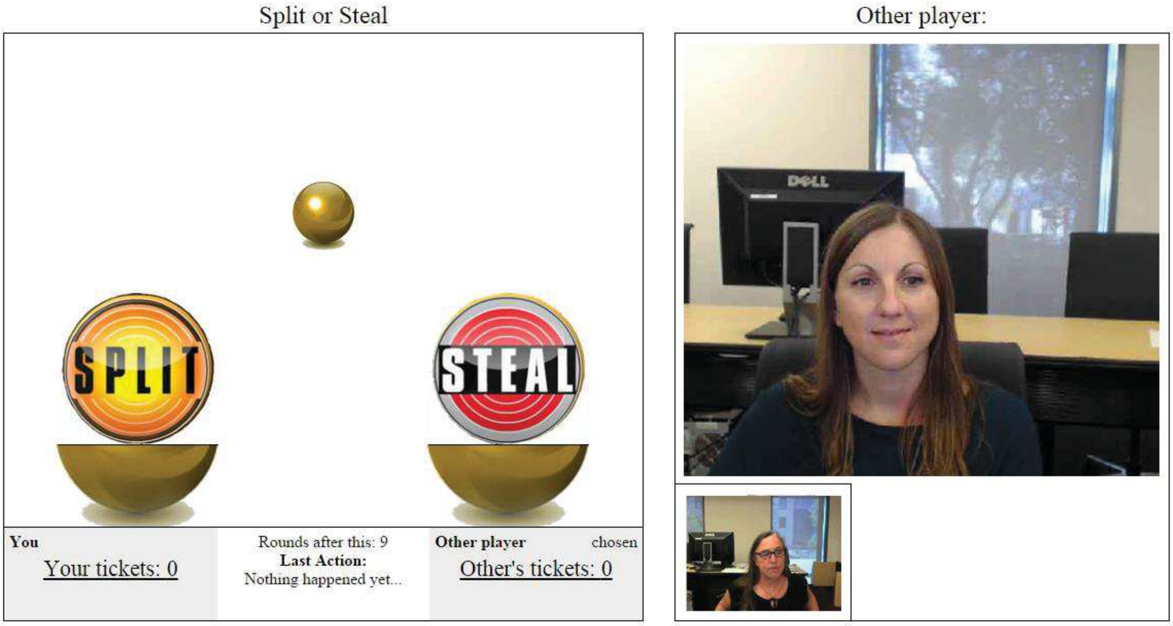

Figure 2. Screenshot of the split/steal game. At the moment the participant can make their choice in the left panel.

3.1.2 Webcam video

While playing the game, participants are recorded by a webcam. The footage of this webcam is sent to a server, as well as to the opponent of the participant. The webcam video is displayed next to the game screen, participants can see both their opponent in a big video as well as their own video in a smaller video on the side. This video is the only way for participants to communicate, as audio is not recorded. Therefore facial expressions are some of the most commonly used ways of communication by participants playing the games.

The videos recorded of participants playing the game are stored on a server as jpeg images, named by their timestamps. In order to use the facial expression recognition software on them, they need to be converted to videos of 30 frames per second. This conversion is done with a few small matlab scripts, as well as the FFmpeg library[3]. Once converted specific features can be extracted from the videos, such as specific expressions and activation of FACS action units.

[3] https://www.ffmpeg.org/

3.1.3 Database

All game events are stored within a MySQL database. The database keeps track of the type of game event, such as the choice of a participant to either cooperate or defect within the prisoner's dilemma. These events are stored by keeping track of all incoming messages through the WebSocket service. Once a game has started each message from the participants will be stored in the database, this way message related to finding users will not be stored.

The time when the event occurred is stored as well, both as the time on the client (by milliseconds) and as time on the database server (rounded to the nearest second). As well as several user identifiers, namely the session ID which is assigned to the client's session. This identifier will be different for each game the user plays, as a new session will be created for each game played. The pair ID combines the session ID of the participant and its opponent. Lastly there is the sender name which is the ID assigned to the user at the start of the study and remains constant throughout the study.

The data in the database can be used to analyze game decision behavior of each participant. The exact time of occurrence of these events can be found because the timing of each decision was being tracked. Using this information an overview of the data was given in a java-based video viewer. As shown in Figure 3 it showed the video of both human participants in a game and showed the specific game events along the videos timeline.

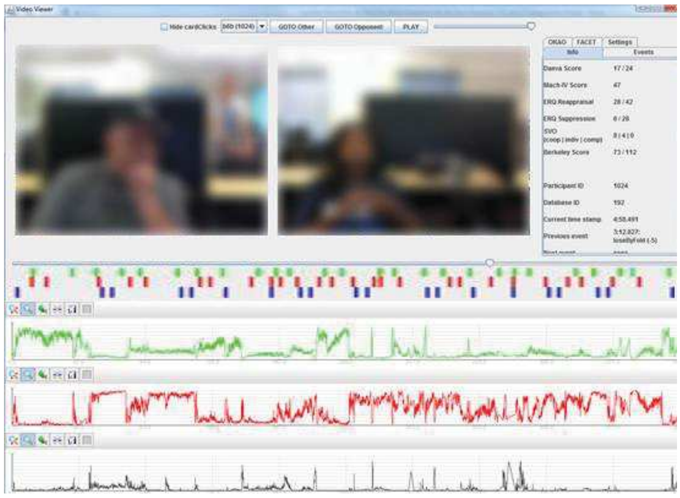

Figure 3. The video viewer, originally build by van Vroonhoven.

3.2 Virtual human

My study focused on the interaction between humans and virtual humans, so the first new addition to the system was the ability to play the iterated prisoner's dilemma against a virtual human. If a participant plays a game against a virtual human, the virtual human will replace the webcam feed of the opponent. The virtual human will be displayed within the "Unity web player", which is downloadable as a plugin for most of the commonly used internet browsers[4].

The virtual human will play the prisoner's dilemma according to a specific strategy defined on the server. It's also possible to create several specific strategies for the virtual human and then assign them based on the specific game state that the user is in or by randomly assigning them. Some examples are: pre-determined, where all the choices are already known at the start of the experiment, random where the

[4] https://unity3d.com/webplayer

agent uses a random move or reactionary where the agent's response is based on the participant's choice. Combinations of these strategies are also possible.

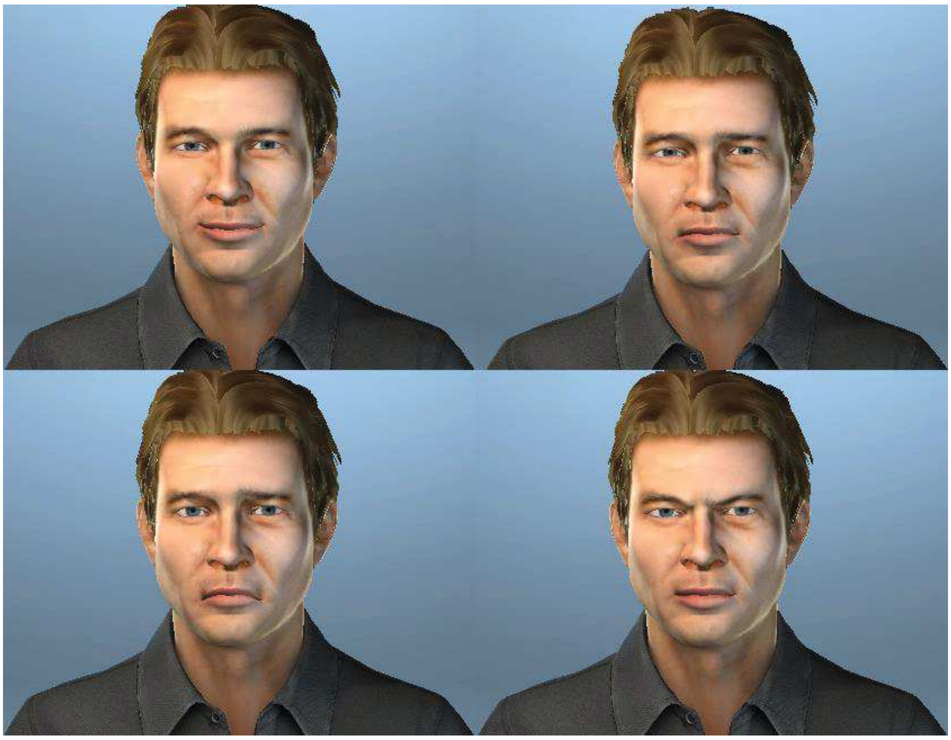

Figure 4. The Virtual Human used in the framework displaying several facial expressions. Clockwise starting in the top left:, Joy, Fear, Anger and Sadness.

The virtual human can use specific facial expressions in order to communicate with the participant, as shown in Figure 4. Additionally in the Ultimatum game it can communicate with text as well. For this communication the virtual human uses specific behavior classes defined on the server. These behavior classes are very similar to the game decision behavior classes, both can be called at any point from the server side.

3.2.1 Limitations

There were several limitations to keep in mind while creating the virtual human for the system. First off, the system was currently only used in lab settings on machines set up by us. One future requirement for the system was to deploy it online, therefore the system would need to be able to run on as many machines as possible.

ICT has a specific system for using their virtual humans within systems, called "Smartbody" [26]. Smartbody allows its users to call animations for the virtual human, such as facial expressions. The functionality of Smartbody could be called by using Behavior Markup Language (BML), an XML-based language. Through BML it is really simple to create scripts for expressed behavior of a VH.

Unfortunately Smartbody was built partially in C++ and as such it was OS dependent, Smartbody would usually work fine on a Windows machine, however not on a system running a UNIX-based OS.

Because of the limitations of Smartbody a different way to use virtual humans within an internet browser had to be considered, in order to allow the system to function on as many systems as possible. Most of ICT's virtual humans ran within Unity, a game engine. In order to process and run animations Smartbody was used within Unity by ICT, however Unity also has its own animation system called Mecanim. Mecanim has only recently been introduced within Unity and as such Mecanim has far less functionality than Smartbody. It does not support BML scripting and its animation system is not as accurate and configurable as Smartbody is.

Despite these limitations, Mecanim sufficed for the system that I was building. ICT has a large library of animations for their virtual humans, most of the facial expression animations are Action Units from the Facial Action Coding System (FACS) of Ekman and Friesen [27]. These animations were stored in files using the FBX format and were directly used by Smartbody to display the animations specified in the BML scripts.

Although Mecanim could not directly use these FBX files, it was possible to import the animations into Unity and convert them to a Mecanim friendly format. This importing was done partially automatically, however it required some fine-tuning by hand. By importing the some animations could be applied to any 3-dimensional computer model that is using a standardized "human skeleton rigging". This means that the joints of the model have to be set up in a similar way to that of a real human in order to perform animations. For example, an animation of someone waving requires movement from the elbow and hand joints, by using the standardized rigging the animation can be applied to any humanoid model. However, in order to get realistic animations more than just these general joints will need to be used. This is especially the case for something like facial expression. Most humanoid models only have 'joints' in their face for movement of their mouth and eyes, facial expressions however will also use the cheekbones and eyebrows. It is possible for Unity to automatically generate these additional joints required for facial expression animations, although in this case some fine-tuning of the joints was required in order for the facial expressions to look more realistic and less uncanny.

Most of the Virtual human's animations will be used as a response to a decision the user made. So the scripts used to call these animations will not be too complicated. Mecanim animations can be called through the general Unity interface, as such it is also not necessary to support BML within the system. Instead a method was created within Unity that can run an animation. It is possible to call this method from the web page that the Unity player plugin is running on. A String can be passed to this method as a parameter in order to specify the type of animation.

Participant (client)	Web Socket	VH (server)
Decide split or steal		Decide split or steal *(based on previous result)*
	Client to server: Decision	
		Determine round result *(Store result)*
	Server to Client: Round result	
Client displays result		Show expression *(based on stored result)*
	Server to Client: VH expression	
VH Expression displayed in Unity		

Figure 5. This model shows how and in which order the VH determines its decisions (Tit-for-tat) and expresions.

Figure 5 shows how the server determines the decisions and expressions of the virtual human. The player starts by making a decision. The virtual human will also make his decision at this point, because both players can make their decision at the same time in the split or steal game. For the first round the VH will always decide to split and on the second round it will always steal. After this it switches to the tit-for-tat strategy by making the same decision as the participant did in the previous round, for more specifics on the VH's strategies in the split-or-steal game refer to section 3.2.3. The decision of the player is sent from the client to the server through a WebSocket. Once both players (participant and VH) made their choice the server will determine the round outcome. After sending this outcome the server will also send the behavior for the vh to use (in this case a facial expression). Both the outcome and the behavior will be sent back to the client through the WebSocket service.

The round outcome will be immediately shown within the game panel using JavaScript functions, because the server can communicate with the JavaScript code directly through the WebSocket. In order to display the animations however, the javascript will need to pass the expression to the Unity Web Player. This is done by calling the animation method within Unity and passing a String giving the specific values of the animation. For example:

```
"Anim- Happy,All,1.0,3.0"
```

This String will be parsed within Unity as follows. First off Unity will use the first part of the String (`Anim`) to determine what method to call, in this case it is the animate method. The second part of the string are the variables for the animate method. The first variable (`Happy`) describes the type of animation. Some animation can be done partially, for example an eyebrow raise can be done with both or only one eyebrow. All is the default value and also only value that can be used for the "happy" animation. The last two float numbers determine the strength of the animation (how pronounced the expression will be) and the length in seconds.

3.2.2 Changes to framework

In many ways the virtual human code functions is similar to how the human opponent code works on the server. The virtual human is a subtype of the User class within the system as it inherits from the General User class as shown in Figure 6. As such a virtual human can do the same things within a game as a human participant could. However unlike the human participants, the virtual human has specific code that determines which decision it ends up making. Additionally the virtual human class has additional functionality which allows it to determine its facial expressions. Most of this functionality is within separate policy classes.

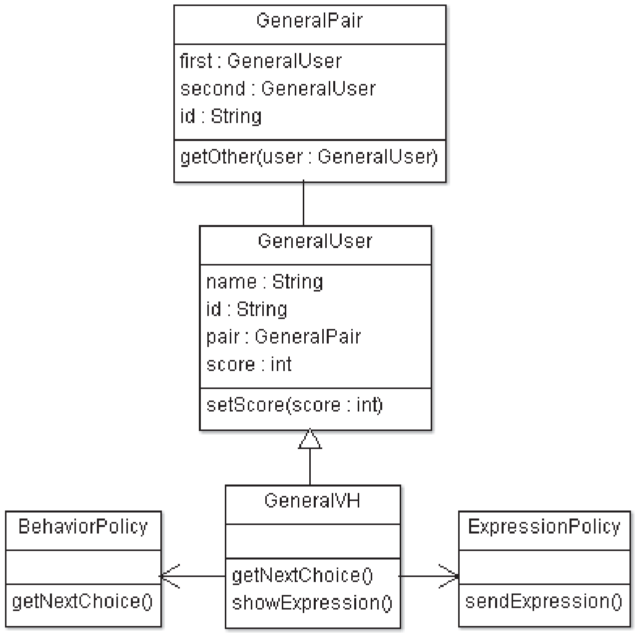

Figure 6. UML of 'User' classes.

The original framework was built specifically for games played by two human participants, much of the old code was written with this in mind. Most of the server's functionality is triggered upon receiving messages from its clients with their decisions. However the virtual human's behavior will be running on the server and will not be using the WebSocket service. Therefore games that are played against virtual human opponents will run slightly different code than those played against real humans. For example, the server will not wait for two WebSocket messages (indicating both users made a decision) before starting a new round. Instead it will start a new round after it received one message from the human user and when the virtual human class has generated a new decision. There were some implications for the system by handling the virtual human like this. The virtual human does not send a message through the WebSocket service, so this decision will not be automatically stored within the database. Therefore some code within the database handler also needed to be changed to ensure all data relating to a game would be properly stored.

Additionally, in human/human games the participants playing against each other will be considered a pair identified by both of their session IDs. While against virtual human opponents this will just be the human participant's session ID, because the virtual human does not exist on a client it does not have a session ID. Both the decisions of the human participant and of the virtual human they were playing against would be stored under the same user like this, therefore the virtual human session ID was altered by having all their session IDs start with the letters "vh".

With these changes to the framework it is now fairly simple to add additional virtual humans to the game, for example you could add a virtual human that behaves differently within the Split or Steal game or create a virtual human for a completely different game. Because all games within the framework rely on participants making decisions, the same general code can be used to define the agent's behavior.

3.2.3 The virtual human in the Split or Steal game

As mentioned before the virtual human can use specific behavior for each game. The virtual human was first implemented within the already existing split or steal game. Within this section I will describe the behavior used by the virtual human in this game. This behavior can be split up in two specific behaviors: The game behavior which describes the way the virtual human plays the game, that is to say how the VH makes its decisions during the game. The expression behavior describes the way the virtual human will display facial expressions while playing the game.

The goal of the virtual human within the split or steal game is to evoke cooperative behavior from the participant. By doing this it will be possible to compare the cooperative behavior the participants show against other humans, with the hypothesis being that people will show more cooperation against humans. This behavior was based on a previous study by de Melo et al. [24], they did several studies with virtual agents displaying different behavior and looked at which agent was cooperated more with by participants.

Figure 7. The virtual human showing a neutral expression in the Split or Steal game

For the game behavior, the virtual human uses a tit-for-tat strategy. This means the virtual human will use the same decision that the participant made in the previous round. Thus, if a participant cooperates, the virtual human will do so in the next round. By doing this the virtual human rewards the participant for cooperative behavior, because it will reciprocate this behavior in the following round. However if a participant defects, the virtual human will punish the participant by defecting as well on the next round.

The virtual human starts this tit-for-tat behavior on the third round of the game, on the first round he will always cooperate and on the second round he will always defect. This set order on the first two rounds is to ensure that every participants will experience both the cooperate and defect decisions of the virtual human.

The expression behavior of the virtual human is designed to elicit cooperative behavior. In the study by de Melo et al. [24] the agent displayed specific facial expressions to encourage the participant to cooperate. When both the participant and the virtual human cooperate this is the best possible outcome and as such the virtual human shows a joyful expression. For any other outcome the virtual human shows more negative expressions as these are not cooperative outcomes.

Table 1. The virtual human's responses to the specific outcomes of rounds in the Split/Steal game

		Virtual human	
		Cooperate (Split)	Defect (Steal)
Participant	Cooperate (Split)	Joy	Fear
	Defect (Steal)	Anger	Sadness

3.3 ULTIMATUM GAME

The Ultimatum game was the first game added to the system after virtual humans were added to it. The game was created with virtual humans in mind. The overall course of an Ultimatum game, as was explained in section 2.1.2, is that one player (in this case the VH) gives the other player a proposal. The other player can then either accept or reject this proposal. The Ultimatum game created for this study was based on a previous study by Mell et al. [6]. The game used in the study was considered difficult to understand by many of the participants.

3.3.1 Original version

Within Mell's game the participant had to play against a virtual agent, represented by a static picture within the game. The game was a playing field made out of several different colored squares. One of these squares was the goal that the player had to try to reach, two other squares contained the position of the participant and the agent they were playing against. In order to move from their position to the goal the participant and virtual agent needed the corresponding colored tiles to the squares they would have to cross in order to reach the goal. Both the participant and the agent have a set of colored tiles at the start of the game, during the game the agent will propose to exchange specific tiles so that either the participant or the agent can make it to the goal. The participant can then decide to either accept or reject this proposal, after doing so the participant could move to the square closest to the goal. The closer the participant came to the goal, the more points he would receive. In the final round of the game the participant would get the opportunity to divide the tiles instead of the agent.

There were different types of agents within the game, some would give favorable proposals for either the participant or the agent, while other types of the agent would only give proposals favorable to themselves.

Additionally the agent was capable of speaking to participants through text. Some of the agents would ask for favors when they would give a proposal that was unfavorable to the participant, by telling the participant that they would repay them. Other agents would not ask for these favors.

Based on this, the study had a two-by-two design, agents would either occasionally give favorable proposals or none at all and agents would either ask for favors or not. The goal of the study was to see whether there would be difference in behavior of participants depending on the type of agent they played with.

3.3.2 New version

The new version of the Ultimatum game had a similar setup as the previous study. The game used the same two-by-two design and as such had four different agent types, however instead of a static agent the participants played with a virtual human. Besides sending text messages, the virtual human will also show either a smile when the participant accepts the offer or a frown if the participant does not. The text messages were defined in a separate new behavior class on the server.

The game itself has also been simplified, instead of reaching a goal on a game board, the new goal is to gather as many tickets as possible. Each round of the game there will be 20 tickets to divide, the tickets are usually divided by the virtual human, who once again can give either good and bad offers or only bad offers. For example a division of 15 tickets for the virtual human and 5 for the participants is a bad offer, while 15 for the participant and 5 for the virtual human is considered a good offer. In the case of the agent giving good and bad offers, this agent will always start with a bad offer then follow it up with a good one and keep alternating like this for the duration of the game.

Because the game is now simply about dividing items, it is much easier to see for the participants when an offer is good or bad. One additional rule to the game is that there are different types of tickets that are not worth the same for both the player and the virtual human. There are tickets with a picture of an apple on them, these are worth more to the virtual human than to the participant, while the tickets with a picture of an orange on it is worth more to the participant. These rounds alternate each other. If the virtual human would give a good offer, it would be on rounds with the orange tickets. As such the bad offer would always be given on rounds with apple tickets, this is also the round where a favor-seeking agent would ask the player for a favor. This favor would then be repaid on the next round, if the agent was set to return favors. Just like the previous study the participant would also get the chance to divide tickets. This occurs during a special round with tickets with bananas on them, these tickets are worth just as much to both the participant and the virtual human.

3.4 INVESTMENT GAME

The investment game was built after running the VH study with virtual humans at ICT. This game was similarly to the Ultimatum game. For one the game was based on a game used in a previous study, this study was done by Shore and Heerey [28], in this study participants played an investment game in a questionnaire environment. Just like the Ultimatum game, participants played the game on their own against an agent. The previous study was done mostly using pictures, the new version improves upon this by using an embodied virtual human. The other aspects of the study remain mostly the same in this new version.

The investment game is played over several rounds. At the start of a round the participant will receive $10, they then get the opportunity to invest this money by sending it to a trustee. Within the new Investment game, a virtual human will take the role of this trustee. The trustee will increase the amount of money invested threefold, after increasing the money the trustee will return a part of the money. This return can either be fair, when the participant receives more than they invested, or unfair when they receive less. After this the next round starts, the results of the previous rounds are not carried over, all the money will disappear and the participants receives a new $10.

The goal of the investment study is to discover whether expressions of guilt will make the virtual human be considered more trustworthy to the participant. The general flow of the game for all participants is as follows. Before starting the game participants would get a set of instructions explaining that the virtual human was controlled by a student on campus. After this participants played 10 rounds against an agent, in all rounds the participants had the opportunity to make an investment which the agent would then respond to by returning a portion of the matured investment.

The study used a 2x2x2 design, there were 2 types of instructions, 2 types of expressions that the agent could use and the agent acted either consistent or inconsistent. First off before starting the game participants would get different instructions, in one of the instructions it was said that the emotion display on the virtual human was done strategically by the student while for the other it was said to be spontaneous. Secondly, the agent would either show a guilt expression after specific turns or smile after specific turns when the return was unfair. The points during which an agent shows an expression were at a set points during the game. Then lastly, based on this expression the agent would either act 'congruent' or 'incongruent' after showing this expression. For example a congruent smiling agent would give a bad return after smiling, as the agent appears to feel good about the previous bad offer. The guilty agent would do exactly the opposite as he feels remorse about his bad offer. The incongruent agent does the exact opposite, so an incongruent smiling agent would give a good return after smiling while an incongruent guilty agent gives a bad return after showing guilt.

There were four types of agents, just like in the Ultimatum game. Unlike the Ultimatum game, participants would not get the chance to play the role of the virtual human in this game. Lastly participants would play the Investment game twice, the expression shown by the virtual human was a between-subjects factor. Therefore participants played a game against both the guilty and smiling trustee, because participants had to believe they were not playing the game against the same opponent, the appearance of the virtual human was slightly different between both games.

4 VH STUDY

In this first study participants played two games, the iterated prisoner's dilemma and the Ultimatum game.

4.1 EXPERIMENTAL SETUP

This study was held in a lab setting. A total of 113 participants were in the study, of which 56 were female. Participants where gathered through an advertisement on craigslist and received $30 for their participation. During the study participants played two different games using the Web Gym framework. Participants played both games on the same machine, after finishing the first game and filling out a short questionnaire on their experiences, the next game would launch right away.

Before playing these games however all participants received a set of instructions that they had to read. This was the emotion regulation manipulation, half of the participants received a form with the actual manipulation of grecucci [4] whereas the other half received a form that described the control condition. The control condition was to simply interpret the expressions of their opponents as they would usually do. The control group received this form in order not to raise suspicion within the emotion regulated group. After reading the form, participants were allowed to introduce themselves to each other during a quick introduction session with the experiment. This session was led by the supervisor of the experiment. The goal of this introduction was to ensure people knew each other and they would not be playing the games against complete strangers.

After this introduction a maximum of five participants were told to take place behind one of several computers. Some plates were placed around these machines, to ensure participants could not see each other directly while playing the game. Participants were also not allowed to speak during the study. The only way for the participants to communicate was through the webcam that was installed on the computers, this webcam only recorded video, not audio. Therefore the only way for participants to communicate with their opponent was through the use of facial expressions, as the group playing against human opponents could see the video from their opponent's webcam on their screens. Participants playing against the virtual human instead saw the virtual human as their opponent.

During the study the first game the participants played was the split/steal game. This was played against either other humans or against a virtual human for ten rounds. The task was based on the one presented by de Melo et al. [25].

The participants played the Split or Steal game against an opponent trying to gather as many "lottery tickets" as possible, the participants were told to gather as many lottery tickets as possible as this would increase their chances of winning the lottery prize of $100. The participants played 10 rounds of the game, after finishing the game they filled out a questionnaire with questions on their own feelings, such as behavior and their opponents, as well as their feelings on the course of the game (see appendix A).

Table 2. Payout matrix for the Split/Steal iterated prisoner's dilemma

		Participant 2	
		Cooperate (Split)	Defect (Steal)
Participant 1	Cooperate (Split)	P1: 5 P2: 5	P1: 0 P2: 10
	Defect (Steal)	P1: 10 P2: 0	P1: 1 P2: 1

The possible outcomes of the player decisions are shown in Table 2. The first number refers to the number of tickets the participants themselves will receive and the second to the number of tickets for the opponent (i.e. if the participant defects and the opponent cooperates, the participant receives ten tickets while the opponent does not get any tickets). As shown in the table, participants will receive the most tickets when they defect upon their opponent, while they get the least number of tickets when they get defected upon themselves. Participants receive more tickets during mutual cooperation than they do during mutual defection. These outcomes satisfy the minimum requirements for the game to be considered a prisoner's dilemma.

The second game played was the Ultimatum game. In the Ultimatum game participants were always matched up against a randomly generated type of virtual human. The types of VHs was based on the two-by-two design of the Ultimatum study. Agents could either use favor language or not and return the favors or not. Just like the Split or Steal game, participants played the Ultimatum for ten rounds.

4.1.1 Analysis of data

The actions of the participants in both games were logged in a database along with the timestamps. Using this data it was possible to determine when decisions were being made, as well as when the reveal and end/begin events occurred for the rounds. The game behavior of the participants was analyzed, such as comparing the number of times a participant would choose to cooperate with a human opponent, compared to a virtual human.

The recorded webcam videos of participants during the study were analyzed using FACET facial expression recognition software[5]. FACET is the commercial version that evolved out of the academic "Computer Expression Recognition Toolbox" (CERT), a system that has reported high accuracy on emotion detection in videos [29]. FACET features include intensities for the basic emotion labels (e.g.

[5] http://www.emotient.com/products/#FACETVision

joy, anger and fear) as well for the higher order labels: “Positive”, “Negative” and “Neutral”. Neutral can be seen as the inverse of expressivity, as it quantifies the absence of showing emotion.

FACET stores intensity values in a comma separated values file of all the labels for each frame in a video, these floating point values range from 0 to 1. In addition to the emotions and high order states. FACET also reports a confidence value for each frame, these confidence values also ranged from 0 to 1. This confidence value is determined by whether FACET can find the main identifiers of the face such as the eyes and mouth. Videos with a low average confidence value were therefore discarded, because of their unreliability. Low confidence values would occur in the case when a face is not completely in frame or if a participant was wearing glasses or other distracting accessories.

During the analysis the focus was on FACET’s higher order values, rather than the emotion values. This was done because the higher order values gave us a more complete picture of the overall feelings of the participants and simplified the analysis. For example, the values for occurrences of the anger emotion could be used to see whether people were enjoying the games, however this would miss the occurrences when the participants displayed fear or contempt during the game. By using the high order label of negative this issue can be avoided. It was possible to obtain the expressions participant showed during specific game events by using the logged the time that was stored in the database. We could for example look at the expression shown by participants when they and their opponents were both cooperating.

Videos with high rate of missing frames were automatically discarded from the analysis. Logging of the game events allowed for automatic event-based behavior encoding as well as automatic segregation of the signals on the game period from the overall recording.

4.2 Results

I will discuss both the results of the analysis on the game behavior of participants and of their expressed behavior within the Split or Steal game in this section. During the study 23 participants played against the virtual human, while the remaining 90 participants played the game against each other in the human condition.

4.2.1 Game behavior

The plots in Figure 8 show how often participants chose to cooperate in both the human and virtual human conditions during the 10 rounds of the game. Participants chose to cooperate more commonly with human opponents than with virtual humans.

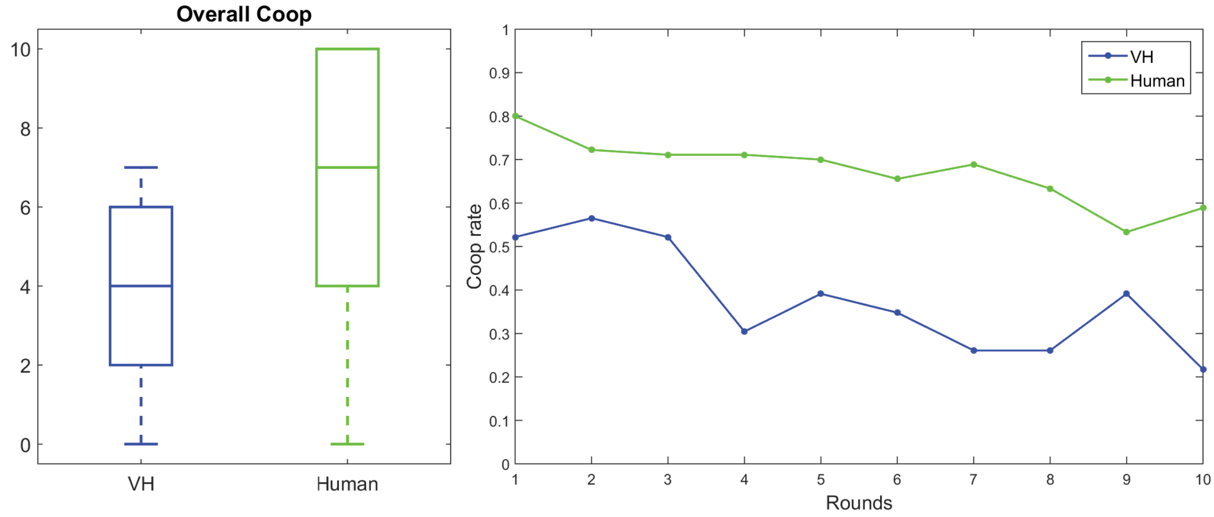

Figure 8. The left boxplot showing the overall cooperation rate for participants playing either a VH or human opponent, the plot on the right displays the cooperation rate per round

An independent T-test was performed on this data, which showed that the difference in overall cooperation between the human and the VH conditions was indeed significant. Participants cooperated with human opponents on average on 6.74 turns (SD=3.14), while only on 3.78 turns (SD=2.19) with virtual human opponents, resulting in the following t-test result: $t(111)=4.26$, $p<0.001$. The plot of the overall cooperation rate in Figure 8 furthermore shows that on average the participants playing a human opponent will always have a higher probability of choosing to cooperate than those that played against the virtual agent. As a second measure of this, the number of times participants chose to defect is also significantly different between the Human (M=3.26 turns, SD=3.14) and the virtual human (M=6.22 turns, SD=2.19) groups; $t(111)=4.26$, $p<0.001$.

The number of times both the participant and their opponents choose to cooperate and as such are in the game state known as joint cooperation, is significantly different as well between both conditions. Participants in the human condition (M=5.16, SD=4.05) are in the joint cooperation state significantly more often than those in the virtual human condition (M=1.70, SD=1.69); $t(111)=4.00$, $p<0.001$. This is also confirmed by the second measure for this, the joint defect state as participants playing humans (M=1.67, SD=2.15) are less in this state than those playing against a VH (M=3.87, SD=2.53); $t(111)=4.23$, $p< 0.001$. This lower amount of being in the joint cooperation state also influenced the final scores, where participants playing against a virtual human scored significantly lower (H: M= 44:10, SD=13.24; VH: M=32.83, SD=10.86), $t(111)=3.77$, $p<0.001$. These results are all in line with the first hypothesis.

Further proof that participants are acting more social against human opponents than against virtual humans can be found in their game decisions: After their opponents defect upon them, the participants are more likely to punish a virtual human opponent (M=3.52 turns, SD=2.29) than a human opponent (M=1.74 turns, SD=2.15); $t(111)=3.49$, $p<0.001$. Please note that the number of turns for this stat is not 10, but 9 as participants cannot respond to their opponents move during the first round. Additionally there is a trend in the data that participants are significantly more likely to exploit (thus being in the "Exploit other" game state) a virtual human opponent (M=2.35 turns, SD=0.93) than a human opponent (M=1.59 turns, SD=1.77); *t*(111)=1.98, *p*=0.050.

One final surprising result is that participants playing against a human opponent employed the Tit-for-Tat strategy significantly more often than those playing against a virtual human. For human opponents participants did this on average 6.59 turns (SD=2.18) while against VHs this was only done for an average of 4.96 turns (SD=1.74); *t*(111)=3.32, *p*=0.001. As the Tit-for-Tat strategy relies on an opponent's previous choice the maximum of Tit-for-tat rounds is 9, not 10. A round is considered Tit-for-Tat when the participant chooses the same choice as his opponent did the previous round. Surprisingly, despite the fact that the virtual human uses a Tit-for-Tat strategy itself, participants did not seem to replicate this strategy during the study.

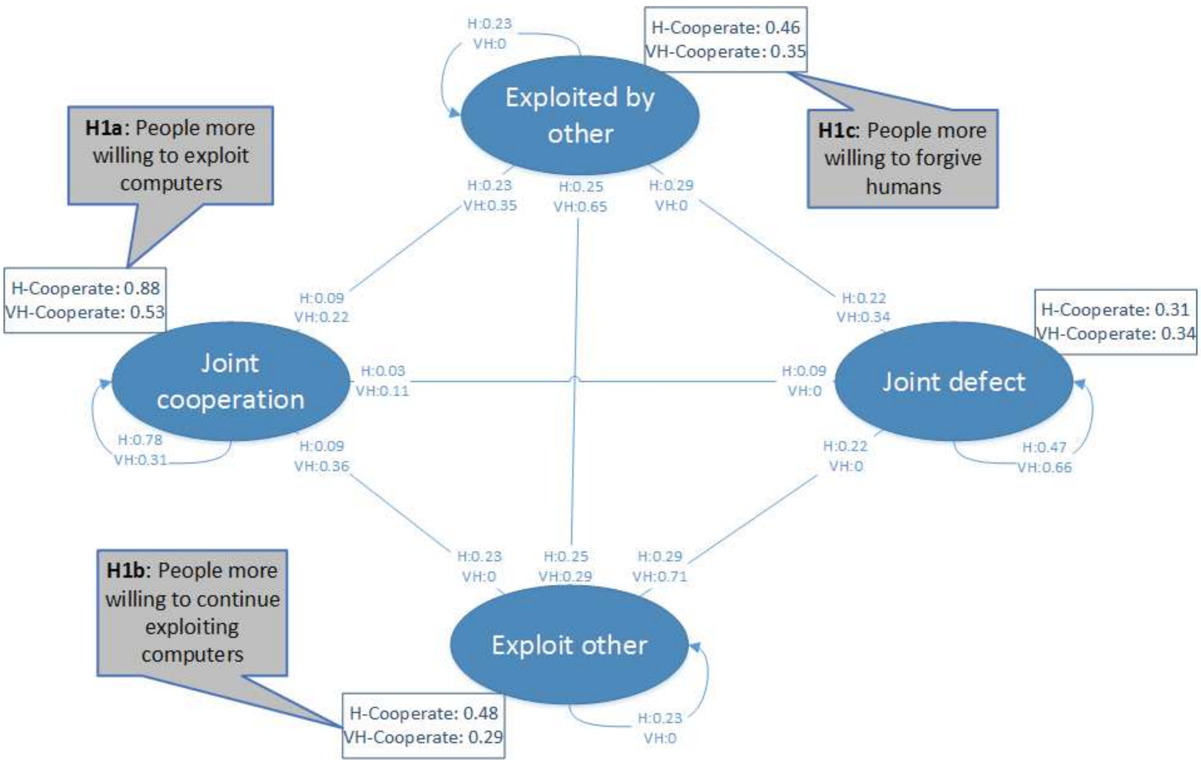

Figure 9. Markov chain of the possible game states. Boxes display the chance a participant would choose to cooperate given in a certain state and support H1.

The Markov Chain in Figure 9 shows that the participants were more willing to exploit virtual human opponents than real humans. When participants are in the mutual cooperation state, the probability of continuing to cooperate is 88% for the human condition, but only 53% for the VH condition. This result is in line with the hypothesis, specifically hypothesis 1a which formulates that people are more willing to exploit computer opponents than real humans.

Participants will also forgive human opponents more easily after being exploited. There is a probability of 46% that a human opponent will be forgiven, whereas for virtual humans this probability is only 35%. This result relates to hypothesis 1c; People are more willing to forgive humans. At the same time participants are more likely to continue defecting on a virtual human opponent after having already exploited them on the previous round. There is a probability of 29% that participants will choose to cooperate again after betraying their virtual human opponent, while this is 48% for real humans. This reflects hypothesis 1b, that people are more willing to continue exploiting computer opponents.

The Markov Chain model shows that overall participants behave more socially against human opponents than they do against virtual human opponents. Participants are more willing to exploit the virtual human, while they are more willing to forgive human opponents if they get defected upon. The only exception to this occurred when participants are in a state of mutual defection, in this case 31% of the participants playing against humans will choose to cooperate, however 34% of the participants playing virtual humans will cooperate.

Participants facing other humans achieve joint cooperation overall more commonly, as overall this group of participants was 52% of the turns in the joint cooperation state. For participants playing against a virtual human this occurred only 17% of the rounds. Participants playing against the virtual human instead were in the mutual defect state for 39% of the games, whereas for the human group this was 17% of the rounds. So despite the fact that the agent was showing expressions that encourage participants to cooperate more often and was using a game behavior to nudge participants into a cooperative direction, participants still achieved mutual cooperation more often with human opponents.

Participants playing against the virtual human also wound up being exploited more commonly, this happened in 21% of the overall outcomes for this group, while participants playing against humans only got betrayed 16% of the rounds. Participants also exploited the virtual human more often than they exploited other humans, they exploited their virtual human opponent 23% of the rounds, while this only happened in 16% of the rounds against other humans.

The self-reported data of the questionnaire that participants filled out after finishing the game also supports the hypotheses. The results of this 7-point Likert scale questionnaire showed that participants considered themselves significantly more cooperative while playing against humans, with a mean score of 5.88 out of 7 (SD=1.72) than against virtual humans with a mean score of 4.57 (SD=1.70), $t(112)=3.27$, $p=0.001$. Participants also reported that they treated humans more fairly (M=5.54, SD=1.85; VH: M=4.30, SD=1.89), $t(112)=2.85$, $p=.001$, further supporting H1.

Similar results were found for the self-reported values on friendliness, honesty and positivity. Participants were playing against a human opponent self-reported as being more friendly (human opponent: M=5.70, SD=1.61; virtual human opponent: (M=4.52, SD=1.81), $t(112)=3.07$, $p=0.003$, more honest (human: M=5.82, SD=1.36; virtual human: M=5.00, SD=1.45), $t(112)=2.56$, $p=.012$ and more positive (human: M=5.98, SD=1.28; virtual human: M=4.96, SD=1.64), $t(112)=3.22$, $p=0.002$. Besides further confirming hypothesis 1, this self-reported data also shows that the data gathered from the game is similar to how participants felt while playing the prisoner's dilemma in van Vroonhoven's study.

4.2.2 Expressed behavior

Besides game behavior the expressed behavior of the participants towards their opponents during the game were analyzed as well. In order to do this, these features were automatically extracted from the webcam video footage using the FACET system. During the analysis the focus was mostly on the "positive", "negative" and "neutral" data that FACET extracted.

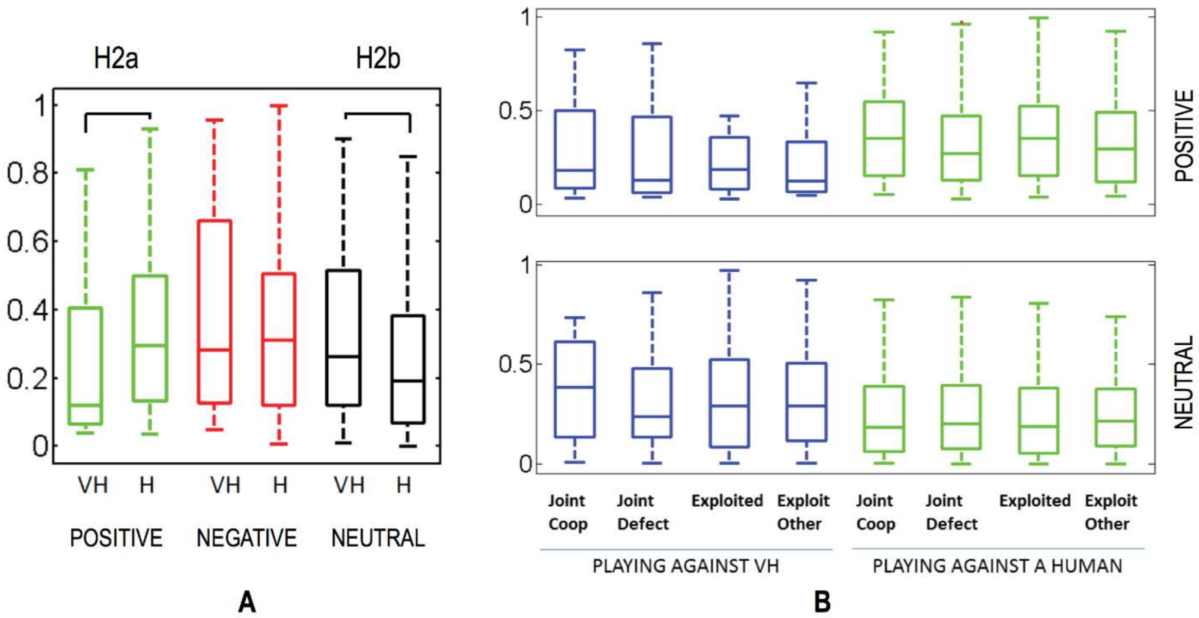

Figure 10. Comparison of displays of expressions when playing with a VH opponent or a human opponent.

The boxplot in Figure 10A shows the average positive, negative and neutral expressivity shown by participants while playing the game against either a human or virtual human opponent. Boxplot A shows that the participants display a trend of higher intensity of positivity against human opponents than they do against virtual human opponents. Similarly to hypothesis 2a, which mentioned that participants would show more joy to human players. Participants also showed a trend of less neutrality when playing against another human than they would when playing against a virtual humans. In hypothesis 2b the prediction was made that participants would show more neutral expression to virtual humans, which is backed up by this trend. A different way to view the neutral statistic is as the inverse of expressivity (the absence of showing positive or negative expressions), as such the data shows people are more expressive against human opponents.

Both these trend in the positive expressivity and neutral expressions were significant at the 0.1 level when using a standard T-test. The boxplot also shows that there was less intensity of negative expressions when interacting with humans, compared with virtual humans, but this was not a significant difference.

The trends in positivity and expressivity combined show that participants express themselves more socially when communicating with a human opponent than they do against virtual human opponents. A possible explanation for this could be that participants are more often in positive game states. As shown in the analysis for game behavior, participants playing against human opponents are in the joint cooperation game state more often. However Figure 10B shows that these trends occur even when the game is broken down by specific game state. The top boxplot shows the positive intensity shown by game state, while the bottom boxplot shows the neutral expressions of participants by game state.

These findings are similar to the reports of de Melo et al. [25] on the self-reported affect during an iterated prisoner's dilemma scenario and support the second hypothesis formulated in this report: That people will display more social expressions to human players compared with VH ones, by showing more positivity to human opponents and less expressivity to virtual humans (i.e. more neutral expression).

The data obtained during this study shows that the displayed expressions are part of an overall behavior of showing less social behavior while playing against a virtual human, as opposed to a human opponent in the game. This is also backed up by the game behavior data. Therefore, it can be argued that facial expressions can be seen as signs of the nature of the dyadic relationship between both participants within the scenarios. When a participant is smiling, they will be doing better in the game as well, as participants cooperate more often with human opponents and display positive expressivity against those opponents as well.

However, facial expressions of emotion are more than a representation of the internal state of a person. Facial expressions are often used as communicative signals to coordinate social interactions between people and have an interpersonal function. Perhaps participants expect that they cannot coordinate with a virtual human as well as they can with a real human opponent and this is what leads to communicating less or less facial expression.

5 STILLS STUDY

One of the main results found in the VH study was an overall lower amount of expressivity of participants while playing the game against a virtual human as compared to a human opponent. During the stills study this was investigated by further dividing the participants that played against a human opponent in two distinct groups. One group would play the game like in the VH study, the other group however would play the game without video footage. Instead participants would only see a still image of their opponents.

The goal was to compare if there was a difference in how participants would interact with a still image of an opponent compared to a video. If this were to be the case, one possible explanation for participants not displaying as much social behavior against a virtual human is because the behavior of the virtual human is not social enough.

5.1 EXPERIMENTAL SETUP

The setup of the stills study was similar to the VH study. Similarly to the VH study, participants would play two types of games in the Web Gym framework. Participants played the Split or Steal game again, however instead of the Ultimatum game, this time participants would play the Investment game (see section 3.4). 118 participants took part in the study, these participants were gathered through a craigslist advertisement and would receive $30 for participating. The study was held in the same laboratory as the VH study.

Participants still would introduce themselves to each other during a quick introduction session, after which the participants would start the study. The computers were set up in a similar way to the VH study and participants were discouraged from speaking during the study. Five participants at most would participate in a study at the same time. Participants would gather lottery tickets during the Split or Steal game again for a chance at the lottery prize of $100. At the end of all games participants would fill out questionnaires on their feelings and behavior during the games.

The Split or Steal game was played exactly the same as it was during the VH study, except that there was no emotion regulation manipulation this time. Instead half of the participants that played against a human opponent would not see video footage of their opponents. If a game was played in this “stills” condition, both participant got the opportunity to take a picture of themselves, once both participants were satisfied with the picture of themselves they could start the game. Instead of video their opponents would only see the picture while playing the game.

The web gym framework had to be altered in order to play the Split or Steal game with still images instead of video. Participants had to take a picture of themselves using their webcam before starting the game, so a new menu was created for this. After reading the instructions of the split/steal game, participants would have to take a picture of themselves using the webcam. This picture would be immediately displayed to the participant after taking it, after this they could either start the game with this picture or retake the picture. Participants could retake this picture as often as they wanted, in order to allow them to get a picture they were satisfied with. This image would then take the place of their video feed. The participant would see the image of their opponent in the video feed and a small version of their own picture in the bottom corner of their opponent's image. Despite the fact that the video was not displayed to participants in the still condition, it still was recorded and could be used in the facial expression analysis using FACET.

The final image the participant took and therefore the one they used during the game was stored on a server for further analysis. A participant in this "still" condition would always be matched up against another participant in this condition. Despite the fact that it was no longer shown to the participants, their video footage was still stored on this server for analysis as well.

The other game participants would play during the study was the Investment game. Participants played through this game twice, once against a virtual human that would smile occasionally and once against a VH displaying a guilt expression instead. The participants would play ten rounds against both virtual humans, in all of the rounds the virtual human took the role of the trustee and the participant that of the investor.

Participants received a manipulation within the instructions for this study as well, however it was a different instruction from the emotion regulation one in the VH study. These instructions were supposed to manipulate the beliefs of participants on the nature of the virtual human during the Investment game, namely whether the VH was intentionally influencing the participant with its facial expressions, while the other half were told that was not the case. This manipulation was for the Investment game and not for the Split or Steal game. For the Split or Steal game participants were simply told that they would play against a virtual human and no further specifics were given. The virtual human in the Split or Steal game also looked differently from the one in the Investment game to emphasize that this was a different type of VH.

5.2 Results

Of the 118 people participating in the stills study, 92 played against human opponents and 26 against a virtual human. The human opponent group was further divided in 46 participants playing with video and 46 with still images. 26 participants played the game against a virtual human opponent.

5.2.1 Video and stills analysis

The first analysis done between the video and still conditions was on expressivity, as the manipulation within the conditions changed the way participants would be able to use their expressions for communication. FACET display expressivity in a scale of 0 to 1, with 1 being the maximum amount of expressivity.

For positive expressions over the entire game, participants with video scored 0.38 (SD=0.24), while participants with video only scored 0.33 (SD=0.21). Participants playing against a VH scored the lowest with a score of only 0.24 (SD=0.19). For neutral expressivity the opposite of the positivity results were found, participants playing against VHs scored the highest with 0.34 (SD=0.28), then participants within the still condition with 0.30 (SD=0.20) and participants with video scored the lowest, 0.26 (SD=0.24). For negativity the participants playing against the VH once again scored the highest with 0.37 (SD=0.29). However the participants in the video condition surprisingly displayed slightly more negativity with 0.31 (SD=0.26) during the entire game than those in the still condition who scored 0.30 (SD=0.26).

An ANOVA was done on all three groups in order to find whether there were significant differences between any of the groups No significant differences were found for negative and neutral expressivity, however a significant difference between these groups was found for positive expressivity $f(2)=3.30$, $p=0.041$. A Tukey post-hoc test showed that the difference in positivity between the video and virtual human conditions was significant ($p=0.031$). Unfortunately no significant differences were found between the video and still conditions.

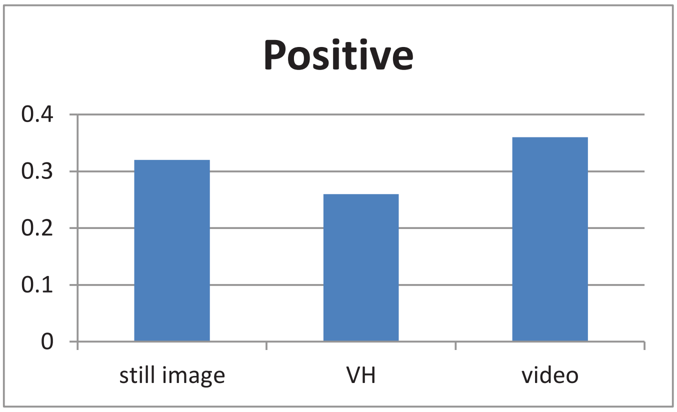

Figure 11. Positive expressivity during the study. The difference between VH and video is significant, however the difference between VH and still image is not.

For the analysis on game behavior some interesting results were found between the video and still conditions. As it turned out, participants in the video condition cooperated significantly more often on a 0.05 level with an average of 6.98 turns (SD=2.74) than those within the still condition (5.63 turns, SD=3.46); $t(90)=2.07$, $p=0.041$. As shown in Figure 12 this higher probability of cooperating occurs in every single round of the game. However there was no significant difference in number of time participants would choose to defect between the groups.

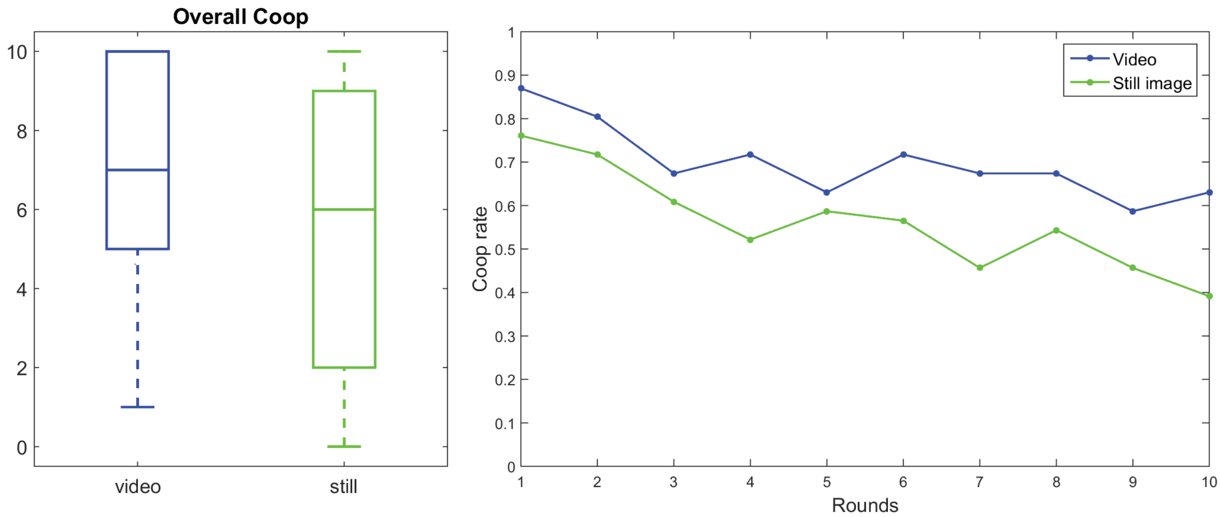

Figure 12. The left boxplot showing the overall cooperation rate for participants playing with video or a still image, the plot on the right displays the cooperation rate per round.

Surprisingly for the specific game states of joint cooperation and joint defection, there was no difference for the joint cooperation state, but there was one for joint defection. Participants in the still condition where in this state 2.65 rounds (SD=2.87), which was more than the 1.43 rounds (SD=1.71) for the participants in the video condition. So for individual choices there was a significant difference in choosing to cooperate, while for the dyadic game state the statistical significant difference was in the joint defect state.

Lastly, participants in the still conditions were more willing to punish an opponent for defecting upon them. Participants in the video condition only did this on 1.52 rounds (SD=1.75), while participants in the still condition did so on 2.50 rounds (SD=2.68). A possible explanation for this could be the fact that participants in the still condition were in the joint defect state more often and as such would have the opportunity to punish more. However it could also be that they were more willing to punish their opponent, as participants could no longer influence their opponent's decisions through the use of facial expressions in this condition.

5.2.2 Comparison with VH study

The Split or Steal game played during the study was the same one as the one played in the VH study. As such the results can be compared between both studies. There are some minor differences however, some participants played against a human opponent without video, seeing instead just a picture of their opponent. The analysis shows that participants not being able to see the video would behave less social to their opponents.

Despite the less social behavior of participants in the still condition, overall participants playing against a human opponent would still cooperate significantly more than those playing against virtual human at a 0.05 level. Participants playing against a human opponent cooperated on average 6.30 rounds out of 10 (SD=3.18), while those playing against a virtual human only did so 4.69 rounds (SD=2.90). $t(116)=2.33$, $p=0.022$. Overall defects was even significant at the 0.01 level, with participants playing against a virtual human choosing to defect on an average of 5.31 rounds (SD=2.90), while those playing against a real human only defected on 3.46 rounds (SD=3.16); $t(116)=2.68$, $p=0.008$.

Similar results to the VH study are also found in the game states participants are in. Participants playing with human opponents are significantly more often in the joint cooperation state (M=4.90, SD=3.87) than those playing against virtual humans (M=2.77, SD=2.89); $t(116)=2.61$, $p=0.010$. Similarly for the joint defect state, participants playing against VHs (M=3.61, SD=2.73) where in this state significantly more often than those playing against humans (M=2.04, SD=2.43); $t(116)=2.84$, $p=0.005$.

Participants are also still more likely to punish opponents for defecting upon them at the 0.05 level. As against VHs (M=3.23, SD=2.66) than against humans (M=2.01, SD=2.32); $t(116)=2.29$, $p=0.24$. At the same 0.05 level participants are also more likely to go Tit-for-Tat against a human opponent (M=6.50, SD=2.20) than against a VH (M=5.42, SD=2.35); $t(116)=2.17$, $p=0.032$.

When combining the datasets of both studies (which gives us 182 participants playing against humans and 49 against virtual humans) many of the significant results are amplified, such as the overall coop rate between human opponents (M=6.52, SD=2.16) and VHs (M=4.27, SD=2.60); $t(229)=4.60$, $p<0.001$.

For expressed behavior the results in study two are somewhat similar to those in study one as well. Although the differences are smaller this time around, perhaps due to the new still image condition, within the human condition. Nonetheless participants display more positivity throughout the games if they are playing against a human than they do against virtual human opponents. Neutral expressivity also seems more common when participants play against a VH compared to real humans, however this is no longer significant at the 0.1 level like it was in study 1. Once again this can be explained by the still condition, if participants cannot see the expressions of their opponents they might show less expressivity themselves.

6 Discussion and conclusion

In these two studies participants played an iterated Prisoner's Dilemma, the Split or Steal game, against either a human or virtual human opponent. In the VH study the social behavior of participants display against either opponents was studied. In the stills study some of the participants that played against a human opponent would not see the video feed, but instead just a picture. It was expected that participants would behave less social against participants if they could not communicate using facial expressions.

6.1 Human and virtual human opponents

During the studies participants showed more social behavior towards other humans than to VHs. This behavior showed in the type of game choices the participant made (game behavior), such as the number of times they chose to forgive or cooperate which they did more with human opponents, or to punish and exploit which was more likely with a virtual human opponent. This result is similar to the first hypothesis that participants would cooperate more with other human players than virtual humans.

A similar result was found for the expressed behavior of the participants. Participants again acted more socially when playing the Prisoner's Dilemma against a human than a virtual human. Participants displayed more cooperative expressions such as joy against a human opponent, while they showed more neutrality against a virtual human opponent. This result was in line with the second hypothesis, that participants would show more cooperative facial expressions to human players compared with virtual humans.

Based on both the game and expressed behavior of participants, an argument can be made that participants behave more socially when facing real human opponents than when facing a virtual human. This is also in line with previous findings of de Melo [17], that although people do treat virtual humans like social entities, they don't treat them equally compared to other humans. This may be because participants do not perceive the same level of emotional expressivity and agency in a virtual human as they do for a real human [20]. In the self-reported data of the questionnaires, participants did indeed report on feeling less connection to a VH opponent (M=2.70, SD=1.49) compared to a human opponent (M=4.46, SD=1.73); $t(112)=4.48$, $p<0.001$. This lower rapport experience within participants could explain why they display more neutral expressions and use less social behavior while interacting with a virtual human than with a human opponent. This is just one of the possible explanations for the behavior of the participants however.

Another possible explanation for the participant's expressed behavior lies in the role that facial expressions play in the Prisoner's Dilemma. Within the study facial expressions are the only means to communicate with the other participant, so if a participant wishes to coordinate with their opponent in order to increase their chances at a good score, using facial expressions to do so is the only way. Therefore if participants do not expect the virtual human to be able to receive these signals as a human would or if they do not believe the virtual human to have the agency to act based on these received signals, this would lead participants to put less effort into channeling these signals and as such show a more neutral expression and less positivity. If participants believe they are in fact less capable to influence the virtual human than a human opponent, this could also explain the significantly lower scores of participants playing against a virtual human.

6.2 VIDEO AND STILL IMAGES

In the stills study the new condition was introduced for participants facing human opponents, where participants would either see the video of their opponent and themselves, or participants would only see a picture that was taken before the game started.

As it turned out, participants would overall show more social behavior against participants if they could see a full video of their opponent. Participants would cooperate more often and be in the joint defect state significantly less if they could see their opponents through the webcam video. While participants in the still image condition were more willing to punish their opponent, by defecting after their opponent defected upon them. There were significant differences in game behavior and participants in the video condition appeared show slightly more positivity and less neutral expressivity. However unfortunately none of these differences in expressed behavior were significant.

Nonetheless the stills study showed that there were some differences between how participants behave depending on how their opponent is shown. The video condition gives participants the opportunity to communicate with their opponent and as such cooperate for better results. The stills condition on the other hand, had participants cooperating less. One possible explanation for this is the fact that the participants no longer had the proper means to communicate. A different explanation however, might be that participants considered their opponent less of a social entity, as described by Blascovich [20]. If this were to be the case, this would mean that if a virtual human's behavior improved, participants might also start thinking of them as more of a social entity and as such display more social and cooperative behavior.

6.3 FUTURE WORK

The goal of this project was to study the difference of displayed behavior against a virtual human compared with humans. While playing the game participants showed significantly less social behavior against virtual human opponents. The exact reason for this change in behavior is future work, there are several explanations that can serve as the outset of new studies.

One way to iterate upon this project is by studying the expectations that participants have of the virtual human and how this influences their social behavior against it. Based on this project's studies, an experiment where the expectations are manipulated can be formulated. For example, the expectation of participants can be manipulation by introducing the virtual human as either an agent controlled by a computer or as an avatar controlled by a human, by stating this within the instructions. If the expectations participants have of the virtual human truly matters, then there might be difference in the behavior shown towards the agent and the avatar condition. Participants might expect a higher possibility of cooperating with a virtual human controlled by a human and therefore show more social and cooperative behavior. Although de Melo has already done a similar study [17], in which participants did indeed respond differently to an avatar than to an AI agent, however by repeating this study using the web gym framework it is also possible to compare the displayed behavior against a real human opponent.

Secondly, according to Blascovich [20], if a virtual human becomes more realistic it, the effect of perceived agency of the virtual human becomes stronger and would start to evoke more social behavior from participants. The virtual human can be improved in several aspects: its game behavior could be made more elaborate and realistic, it's expressed behavior more human-like and the animations and graphics of the virtual human could be improved. Based on Blascovich these improvements would then lead to more social behavior of participants.

During this study a lot of game behavior data was recorded, both for participants playing against a VH and against a human opponent. One might argue that in order for a virtual human to behave more realistic and human-like, this behavior could be based on that of the actual participants in these studies. As shown in Figure 9, it is possible to map the game behavior to a network structure, which could then be used as behavior of a virtual human within the Web Gym framework. However also aspects of the games such as decision time have been recorded in a database and can be used to help the virtual human's behavior become more human-like. The same can be done for the expressed behavior by using the extracted FACET software data. By determining when participants display expressions in relation to specific game events, such as after making a split or steal decision and after the round results are revealed, and using this data to determine the expressed behavior of the virtual human. If more perceived realism truly leads to more social behavior, these behavior models should perform better at evoking cooperation than the models used during this project.

Another way of improving the realism of the virtual human is by improving the quality of its appearance to make it look more realistic. This can be done for both virtual human as a whole or by improving specific animated actions that the VH performs while playing the game. There is an ongoing debate on whether the quality of an agent's behavior or its appearance have the biggest impact on whether an agent is perceived as a social being. For example, Bailenson et al. [30] describe that both depend on each other. According to Bailenson et al. the effect of realism will be minimized if one aspect is less than the other; the agent might seem very realistic but not act like it, or act realistic but not appear realistic at all. However the fact that perceived agency is a large aspect of perceived realism, seems to point to the fact that behavior might be more important, agency is after all based on the behavior of the agent. In which case improving the appearance is not as important as improving the agent's behavior. This too can be studied, by comparing VHs with different types of behaviors and appearances. Additionally this could further be expanded by comparing the influence of this 'perception' (both behavioral and appearance) of the agent and compare it with the 'expectation' participants have using the avatar/AI agent manipulation.

One last topic of future work is the possibility of predicting choices made by participants using the data gathered throughout these studies. This could also be used to make the virtual human more cooperative, if it could predict when a participant wishes to cooperate. By using data such as the facial expressions of the participant, the history of game decisions and the time it takes them to decide, it might be possible to determine what kind of choice a participant is going to make.

6.4 Conclusion

There still appears to be a large difference in how people treat a virtual human compared to a real human. Within scenarios as those presented during the studies of this project, virtual humans were treated with less social behavior than human opponents were. The analysis of game behavior shows that participants cooperate less often with virtual humans than they do with real humans. When participants are in a state of mutual cooperation (i.e. both participants have chosen to cooperate), they are more likely to betray a virtual human by choosing to defect and are also more likely to continue defecting upon the virtual human from this point on. If a participant is betrayed, they are more likely to forgive a human opponent by choosing to continue cooperating themselves. This trend also occurs in the expressed behavior of the participants, as participants show more positive facial expressions against human opponents. Against a virtual human participants show less expressivity overall than against human opponents.

The virtual human was using previously tested behavioral and expression models that were proven to be effective at evoking cooperation between participant and VH. Despite the use of this model, the level of cooperation between two humans was still higher than between a human and a virtual human. Based on the expressivity of participants and data from the questionnaires, participants seemed to enjoy playing against a real human more than with a virtual human.

The stills study showed some trends that communication is important for participants during this study. Participants will show less expressivity when they no longer have a video feed of their opponent and themselves. The still-condition did not just influence the expressed behavior of participants, but also their game decisions. Participants were able to cooperate more often when they could see each other's videos than in the still-conditions. However, participants still showed more social behavior against a human without video than to a virtual human.

These results seem to implicate that while the expressed behavior has some influence, it is not the only aspect that leads to people treating virtual humans differently from real humans. Game decisions could have an impact on this as well. The expectations participants have of their opponents, such as whether it is an actual human (or an avatar controlled by a human) or an AI, might play a role as well.

The results of this project show that virtual humans still have a long way to go before they can evoke truly human-like responses. Based on this study, there are several way to improve and advance the behavior of virtual humans in order to receive more social treatment. By understanding the impact of people's expectations of the virtual human, it might be possible to evoke more cooperative behavior of people with the VH. Data gathered during this study can also be used to improve our understanding on how people act in these kinds of scenarios. This data can in turn be used to enhance the behavior of the virtual human and its realism, which in turn hopefully will lead to it being treated more socially.

References

[1] P. Kenny, A. Hartholt, J. Gratch, W. Swartout, D. Traum, S. Marsella, and D. Piepol, “Building interactive virtual humans for training environments,” in *Proceedings of I/ITSEC*, 2007, vol. 174.

[2] J. J. Gross, “The emerging field of emotion regulation: An integrative review,” *Rev. Gen. Psychol.*, vol. 2, no. 5, pp. 271–299, 1998.

[3] J. van Vroonhoven, “Manipulating Game Decisions with Facial Expressions.” 2015.

[4] A. Grecucci, C. Giorgetta, N. Bonini, and A. G. Sanfey, “Living emotions, avoiding emotions: Behavioral investigation of the regulation of socially driven emotions,” *Front. Psychol.*, vol. 3, no. JAN, pp. 1–14, 2013.

[5] W. Güth, R. Schmittberger, and B. Schwarze, “An experimental analysis of ultimatum bargaining,” *J. Econ. Behav. Organ.*, vol. 3, no. 4, pp. 367–388, 1982.

[6] J. Mell, G. Lucas, and J. Gratch, “An Effective Conversation Tactic for Creating Value over Repeated Negotiations,” *Proc. 14th Int. Conf. Auton. Agents Multiagent Syst.*, pp. 1567–1576, 2015.

[7] W. Poundstone, “Prisoner’s Dilemma: John von Neuman, Game Theory, and the Puzzle of the Bomb.” Doubleday, New York, 1992.

[8] J. Berg, J. Dickhaut, and K. McCabe, “Trust, reciprocity, and social history,” *Games Econ. Behav.*, vol. 10, no. 1, pp. 122–142, 1995.

[9] K. R. Thórisson, “Gandalf: an embodied humanoid capable of real-time multimodal dialogue with people,” in *Agents*, 1997, pp. 536–537.

[10] W. L. Johnson, J. W. Rickel, and J. C. Lester, “Animated pedagogical agents: Face-to-face interaction in interactive learning environments,” *Int. J. Artif. Intell. Educ.*, vol. 11, no. 1, pp. 47–78, 2000.

[11] S. C. Marsella, W. L. Johnson, and C. LaBore, “Interactive pedagogical drama,” in *Proceedings of the fourth international conference on Autonomous agents*, 2000, pp. 301–308.

[12] J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan, and N. Badler, “Creating interactive virtual humans: Some assembly required,” 2002.

[13] C. Becker-Asano and I. Wachsmuth, “Affective computing with primary and secondary emotions in a virtual human,” *Auton. Agent. Multi. Agent. Syst.*, vol. 20, no. 1, pp. 32–49, 2009.

[14] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, “Human computing and machine understanding of human behavior: a survey,” pp. 13–24, 2007.

[15] Z. M. Ruttkay, D. Reidsma, and A. Nijholt, “Human computing, virtual humans and artificial imperfection,” in *Proceedings of the 8th international conference on Multimodal interfaces*, 2006, pp. 179–184.

[16] C. M. De Melo, P. Carnevale, and J. Gratch, “The effect of virtual agents’ emotion displays and appraisals on people's decision making in negotiation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7502 LNAI, pp. 53–66, 2012.

[17] C. M. de Melo, J. Gratch, and P. J. Carnevale, “The Effect of Agency on the Impact of Emotion Expressions on People’s Decision Making,” in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, 2013, pp. 546–551.

[18] B. Reeves and C. Nass, *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press, 1996.

[19] C. Nass and Y. Moon, “Machines and mindlessness: Social responses to computers,” *J. Soc. Issues*, vol. 56, no. 1, pp. 81–103, 2000.

[20] J. Blascovich, “A theoretical model of social influence for increasing the utility of collaborative virtual environments,” in *Proceedings of the 4th international Conference on Collaborative Virtual Environments*, 2002, pp. 25–30.

[21] R. Riedl, P. Mohr, P. Kenning, F. Davis, and H. Heekeren, “Trusting humans and avatars: Behavioral and neural evidence,” 2011.

[22] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher, “Can machines think? Interaction and perspective taking with robots investigated via fMRI,” *PLoS One*, vol. 3, no. 7, p. e2597, 2008.

[23] G. A. van Kleef, C. K. W. de Dreu, and A. S. R. Manstead, “An interpersonal approach to emotion in social decision making: The emotions as social information model,” *Adv. Exp. Soc. Psychol.*, vol. 42, pp. 45–96, 2010.

[24] C. M. de Melo, P. J. Carnevale, S. J. Read, and J. Gratch, “Reading people’s minds from emotion expressions in interdependent decision making,” *J. Pers. Soc. Psychol.*, vol. 106, no. 1, pp. 73–88, 2014.

[25] C. M. de Melo, J. Gratch, and P. J. Carnevale, “The Importance of Cognition and Affect for Artificially Intelligent Decision Makers,” in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[26] M. Thiebaux, S. Marsella, A. N. Marshall, and M. Kallmann, “Smartbody: Behavior realization for embodied conversational agents,” in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, 2008, pp. 151–158.

[27] P. Ekman and W. V Friesen, “Facial action coding system,” 1977.

[28] D. M. Shore and E. A. Heerey, “Do social utility judgments influence attentional processing?,” *Cognition*, vol. 129, no. 1, pp. 114–122, 2013.

[29] M. Bartlett, G. Littlewort, T. Wu, and J. Movellan, "Computer expression recognition toolbox," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, 2008, pp. 1–2.

[30] J. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich, "The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments," *Presence*, vol. 14, no. 4, pp. 379–393, 2005.

APPENDIX A: GAME REVIEW QUESTIONNAIRE

Q1 Please rate the degrees of feelings you have after playing this game.
(0:100) I feel fear
(0:100) I feel joy
(0:100) I feel sadness
(0:100) I feel hope
(0:100) I feel disappointment
(0:100) I feel relief

Q2 How important to you was winning this game?
- ❍ Not at all Important
- ❍ A little Important
- ❍ Neither Important nor Unimportant
- ❍ Very Important
- ❍ Extremely Important

Q3 How much effort did you devote to winning this game?
- ❍ None
- ❍ Some
- ❍ Quite a Bit
- ❍ An Extreme Amount
- ❍ All

Q4 How much personal control did you have over who won the game?
- ❍ None
- ❍ Little
- ❍ Some
- ❍ A Lot

Q5 How satisfied were you with the final score?
- ❍ Very Dissatisfied
- ❍ Dissatisfied
- ❍ Somewhat Dissatisfied
- ❍ Neutral
- ❍ Somewhat Satisfied
- ❍ Satisfied
- ❍ Very Satisfied

Q6 How satisfied do you think the other player is with the final score?
- ❍ Very Dissatisfied
- ❍ Dissatisfied
- ❍ Somewhat Dissatisfied
- ❍ Neutral
- ❍ Somewhat Satisfied
- ❍ Satisfied
- ❍ Very Satisfied

Q7 What was your impression of the other player?

	1	2	3	4	5	6	7
Uncooperative:Cooperative	❍	❍	❍	❍	❍	❍	❍
Unfriendly:Friendly	❍	❍	❍	❍	❍	❍	❍
Reactive:Strategic	❍	❍	❍	❍	❍	❍	❍
Dishonest:Honest	❍	❍	❍	❍	❍	❍	❍
Selfish:Fair	❍	❍	❍	❍	❍	❍	❍
Ineffective:Effective	❍	❍	❍	❍	❍	❍	❍
Poor communicator:Good Communicator	❍	❍	❍	❍	❍	❍	❍
Negative:Positive	❍	❍	❍	❍	❍	❍	❍

Q8 How would you describe your own behavior?

	1	2	3	4	5	6	7
Uncooperative:Cooperative	❍	❍	❍	❍	❍	❍	❍
Unfriendly:Friendly	❍	❍	❍	❍	❍	❍	❍
Reactive:Strategic	❍	❍	❍	❍	❍	❍	❍
Dishonest:Honest	❍	❍	❍	❍	❍	❍	❍
Selfish:Fair	❍	❍	❍	❍	❍	❍	❍
Ineffective:Effective	❍	❍	❍	❍	❍	❍	❍
Poor communicator:Good communicator	❍	❍	❍	❍	❍	❍	❍
Negative:Positive	❍	❍	❍	❍	❍	❍	❍

Q9 How difficult did you find it to get a good score?
- ❍ Very Difficult
- ❍ Difficult
- ❍ Somewhat Difficult
- ❍ Neutral
- ❍ Somewhat Easy
- ❍ Easy
- ❍ Very Easy

Q10 Please consider the extent to which you felt each of the following conditions during play

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
I felt I had a connection with the other player	❍	❍	❍	❍	❍	❍	❍
I think that we understood each other	❍	❍	❍	❍	❍	❍	❍
The other player was warm and caring	❍	❍	❍	❍	❍	❍	❍
The other player was respectful to me	❍	❍	❍	❍	❍	❍	❍
I felt I had no connection with the other player	❍	❍	❍	❍	❍	❍	❍
The other player created a sense of closeness or camaraderie between us	❍	❍	❍	❍	❍	❍	❍
The other player created a sense of distance between us	❍	❍	❍	❍	❍	❍	❍
The other player communicated coldness rather than warmth	❍	❍	❍	❍	❍	❍	❍
I wanted to maintain a sense of distance between us	❍	❍	❍	❍	❍	❍	❍
I tried to create a sense of closeness or camaraderie between us	❍	❍	❍	❍	❍	❍	❍
I tried to communicate coldness rather than warmth	❍	❍	❍	❍	❍	❍	❍

Q11 Did you employ any kind of strategy, and if so, what strategy did you take?

Q12 Was there anything confusing about this game? If so, please describe.

APPENDIX B: FRAMEWORK CODE

The code will be submitted separately to the members of the examination committee in a zip file.

APPENDIX C: PAPER

I worked on the following paper during my final project. It was accepted as a short paper in the Fifteenth International Conference on Intelligent Virtual Agents (IVA), which will be held on August 26-28 2015 in Delft.

Comparing behavior towards humans and virtual humans in a social dilemma

Rens Hoegen, Giota Stratou, Gale M. Lucas, and Jonathan Gratch

University of Southern California, Institute for Creative Technologies
{rhoegen,stratou,lucas,gratch}@ict.usc.edu

Abstract. The difference of shown social behavior towards virtual humans and real humans has been subject to much research. Many of these studies compare virtual humans (VH) that are presented as either virtual agents controlled by a computer or as avatars controlled by real humans. In this study we directly compare VHs with real humans. Participants played an economic game against a computer-controlled VH or a visible human opponent. Decisions made throughout the game were logged, additionally participants' faces were filmed during the study and analyzed with expression recognition software. The analysis of choices showed participants are far more willing to violate social norms with VHs: they are more willing to steal and less willing to forgive. Facial expressions show trends that suggest they are treating VHs less socially. The results highlight, that even in impoverished social interactions, VHs have a long way to go before they can evoke truly human-like responses.

Keywords: Virtual Humans, Social Behavior, Facial Expressions, Decision Making

1 Introduction

Do people treat machines like people? This has been a central concern within the virtual agent and robotics communities, almost since their inception. The answer to this question has more than passing interest. Virtual humans (VH) are increasingly used to teach people how to interact with other people. VHs teach people how to negotiate [4] or how to overcome fear of public speaking [2]. Others have proposed virtual agents or robots as replacements for people in a variety of customer service and even business settings. Following Cliff Nass' early work on the Media Equation [14], it is common to assume that the same social processes arise in both human and VH interaction, and many subsequent studies have reinforced the validity of this assumption (e.g., [17]).

Yet, recent studies emphasize important differences in how people treat machines [9, 6]. Further, there is good reason to believe that studies under-report the differences between humans and artificial partners, as most "direct comparisons" are not as direct as they might seem. The most common method is to manipulate the *mere belief* of who one is interacting with. For example, people interact with a digital character but in one case they believe they are playing a computer program and in the other case they believe a person is driving the agents behavior [9]. While there are good methodological reasons to adopt this experimental design, it also clearly under-represents the differences between human and VH interaction. It is a necessary but insufficient step towards demonstrating equivalence between human and machine interaction.

In this study, we make a direct comparison between the behavior of people interacting with other humans in face-to-face interaction with their behavior when interacting with VHs. We explore this within the context of a standard economic game, the iterated prisoner's dilemma, as this allows for several behavioral measures and allows us to connect our findings with a number of existing studies on social behavior. Prior VH research and robotics research on the prisoner's dilemma manipulated the beliefs of participants as to whether they were playing a real or virtual human, but decisions were always made by a computer (e.g., [7, 10]). However, in this study we compared data between humans that could see each other, but not speak, via webcam, against humans playing an emotionally-expressive VH. In order to determine social behavior we analyzed both the strategy used by participants and their use of facial expressions, by using facial expression recognition software. Based on prior findings on how people treat VHs in this game [7], we hypothesize that people will be more reluctant to show pro-social behavior to a VH, both through their actions and emotional displays. We explain these hypotheses in the next section.

In section 2 we will give an overview of work involving the displayed social behavior against VHs. Section 3 describes the specifics of the iterated prisoner's dilemma game played during the study, as well as the VH that was used. Furthermore information on the analysis of the game behavior and expressed behavior will be given. Section 4 contains the overview of the results of the study for both game and expressed behavior and in section 5 the implications of these results will be discussed.

2 Related work

There are several views on the social behavior people show when interacting with computers. The Media Equation of Reeves and Nass claimed that responses to computers would equal responses to humans when computers incorporate human-like social cues. This was claimed to occur because people develop automatic responses to social cues and thus, unconsciously react automatically to computers in the same way as they do towards other humans [13]. It has been argued that the concept of facial expressions within economic games can serve as automatic elicitors of social behavior [16] and that VHs can exploit these cues [5].

A strong interpretation of the media equation, often articulated by Nass [12], is that responses towards computers are equivalent to human responses when computers incorporate social cues. A more nuanced perspective replaces the "=" in Nass' media equation with a "<". Blascovich [3] argues that social influence will increase based on the perceived realism and "agency" of a virtual agent. Agency refers to the perceived sentience or free will of an agent. This view is supported by a study of de Melo et al. [7], in two experiments agency was manipulated by comparing VHs that were either agents (i.e. controlled by computers) with avatars (i.e. controlled by humans). These experiments showed that people cooperated more with VHs which showed specific facial displays, however these displays only scored significantly different for the avatar condition, thus showing the difference in social behavior people display while playing against humans or VH. Riedl et al. [15] have done a study where participants played against humans and avatars. Their results showed that people display similar trust behavior between humans and avatars, however through neuroimaging they showed that there are different responses in the brain between human and avatar opponents. In a study by Krach et al. [10] humans played an iterated prisoner's dilemma, against both computers, robots and humans. Their results showed that humans in fact experienced more fun and competition in the interaction with increasing human-like features of their partners.

Our current study builds on the findings of de Melo et al. [8]. In their study, people played an iterated prisoners dilemma with a VH that played tit-for-tat and expressed specific emotions. In one condition, participants believed the agents choices and emotions were selected by another participant. In the other, they believed they were generated by a computer programmed to behave like a human. In either case, players could send emotional expressions to the other player along with their choice in the game. The tit-for-tat behavior and the pattern of emotional expressions were both chosen to maximize the amount of cooperation shown by participants. Nonetheless, participants made less cooperative choices and sent fewer positive and more neutral expressions when they believed they were playing a computer opponent. Based on these findings, we make the following hypotheses:

> **H1**: *Participants will cooperate significantly more with other human players than VHs. More specifically, we predict people will be (H1a) more willing to try to exploit a VH, (H1b) more willing to persist in exploiting a VH, and (H1c) more willing to forgive humans following exploitation.*

> **H2**: *Participants will show more cooperative facial expressions to human players compared with VHs. Specifically, we predict people will (H2a) show more joy to human players and (H2b) show more neutral expressions to VH.*

3 Experimental Setup

For this study, participants played an iterated prisoner's dilemma game against either other humans or a VH. The study used a 2-cell design, a total of 113 participants (56 female) participated in this study. 23 participants played against the VH, the remaining 90 participants played the game against each other in the human condition. No specific information was given on the VH, participants were simply told that they would play the game against either a human or a

		Opponent	
		Cooperate	Defect
Participant	Cooperate	5	0
	Defect	10	1

		Virtual human	
		Cooperate	Defect
Participant	Cooperate	Joy	Fear
	Defect	Anger	Sadness

Table 1. Left: Number of tickets the participant receives per outcome. Right: VH responses to outcomes.

virtual human. The task was based on the one presented by de Melo et al. [7]. Participants played 10 rounds of the game and the possible outcomes of the player decisions are shown in Table 1.

The game interface, shown in Figure 1, displays the game on one side of the screen and the opponent on the other side. The participants chose whether to "split" the tickets or try to "steal" them, corresponding to the cooperate and defect options.

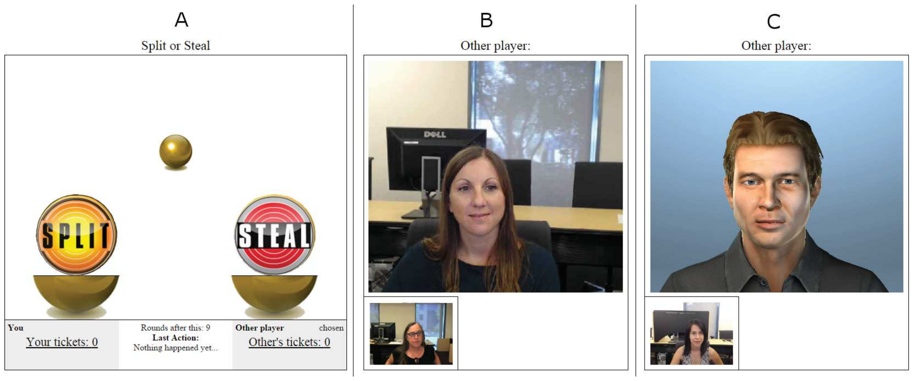

Fig. 1. Screenshot of the split/steal game. Panel A shows the game at the moment the participant can make their choice. Panel B will only be shown to participants in the human condition, panel C only for the VH condition.

Figure 1 shows both the human and VH condition of the game. The experiment was performed in a lab setting, with a maximum of five participants playing the game on computers. Participants were not allowed to speak during the study. The group playing against human opponents could see the video from their opponents' webcam on their screens. Participants playing against the VH would instead see the VH display within the web browser using the Unity web player plugin.[1] The VH used a tit-for-tat strategy during the game, similar to the study by de Melo et al [7]. The agent used this strategy for the entire game with the exception of the first and second round, on the first round the agent would always cooperate with the participant, whereas on the second round the

[1] `https://unity3d.com/webplayer`

VH would always defect. Table 1 shows the facial expressions feedback of the VH on the outcome of a round. These expressions are based on the expressions of virtual agents tested in a study by de Melo et al. [5] and were found to perform the best at eliciting cooperation. The actions of the participants were logged in a database along with the timestamps. Using this data we could infer when decisions were made, when the reveal and when rounds began or ended.

Participant videos from the webcams were automatically analyzed using FACET facial expression recognition software.[2] FACET features include intensities for the basic emotion labels as well for overall sentiment labels: "POSITIVE", "NEGATIVE" and "NEUTRAL". FACET is a commercial software for expression recognition that evolved from an academic version, the "Computer Expression Recognition Toolbox" (CERT) [1] and reports high accuracy on emotion recognition labels on known datasets [11]. Videos with high rate of missing frames were automatically discarded from the analysis. Logging of the game events allowed for automatic event-based behavior encoding as well as automatic segregation of the signals on the game period from the overall recording.

4 Results

This section describes the results of our study. Section 4.1 shows our findings on H1, section 4.2 the findings on H2.

4.1 Game Behaviors

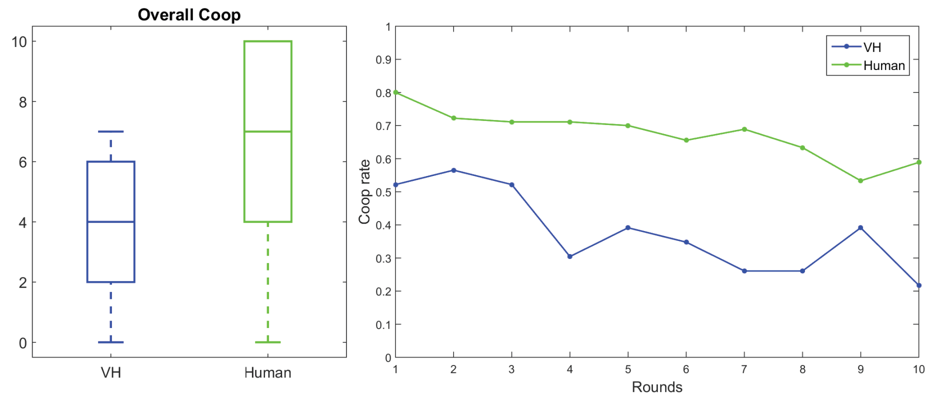

Fig. 2. The left boxplot showing the overall cooperation rate for participants playing either a VH or human opponent, the plot on the right displays the cooperation rate per round

[2] http://www.emotient.com/products/#FACETVision

The plots in Figure 2 show how often participants chose to cooperate in both conditions. We performed an independent T-test on this data, which showed that there was a significant difference in overall cooperation between the human (M=6.77, SD=3.17) and the VH conditions (M=3.64, SD=2.13); $t(108)=4.39$, $p<0.001$. The number of times participants perform mutual cooperation, when in the human condition (M=5.19, SD=4.07) are in this game state significantly more often than those in the VH condition (M=1.59, SD=1.65); $t(108)=4.06$, $p<0.001$, whereas the opposite is true for the mutual defect state (Human: M=1.69, SD=2.17; VH: M=4.05, SD=2.44); $t(108)=4.44$, $p<0.001$.

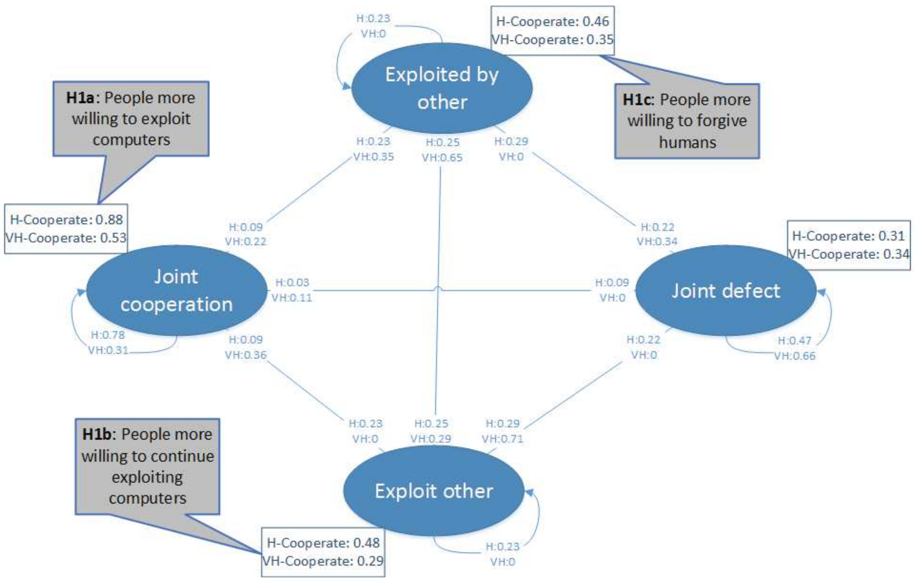

Fig. 3. Markov chain of the possible game states. Boxes display the chance a participant would choose to cooperate given in a certain state and support H1.

The Markov Chain in figure 3 shows that participants are generally more likely to exploit VH opponents than real humans (H1a). When participants are in a mutual cooperation state, the probability of continuing to cooperate is 88% for the human condition, but only 53% for the VH condition. Participants will forgive human opponents more easily after being exploited (H1c), with a 46% probability, whereas for VH it is only 35%. Similarly, participants are more likely to continue defecting on a VH opponent after having already exploited them once (H1b), with a probability of 29% participants will choose to cooperate again after betraying their VH opponent, while this is 48% for real humans.

Participants facing other humans are overall more likely to achieve joint cooperation with a probability of 52% while only 17% for the VH group. Participants playing against the VH instead had a chance of 39% to reach mutual defect, whereas for the human group 17%.

The self-report questionnaire data also supports the same hypothesis. On a 7-point Likert scale people considered themselves significantly more cooperative while playing against humans (M=5.88, SD=1.72) than against VHs (M=4.57, SD=1.70), $t(112)=3.27$, $p=0.001$. Participants also self-reported that they were more fair against humans (M=5.54, SD=1.85; VH: M=4.30, SD=1.89), $t(112)=2.85$, $p=.001$, further supporting H1. We found similar results in the self-reported data on friendliness, honesty and positivity.

4.2 Expressed Behaviors

As a secondary aspect of the behavior towards the game opponent we examined the participant displays of emotion during the game. For this purpose we used the automatically extracted measures mentioned in Section 3 and looked mainly at the intensities of the summary labels: "POSITIVE", "NEGATIVE" and "NEUTRAL".

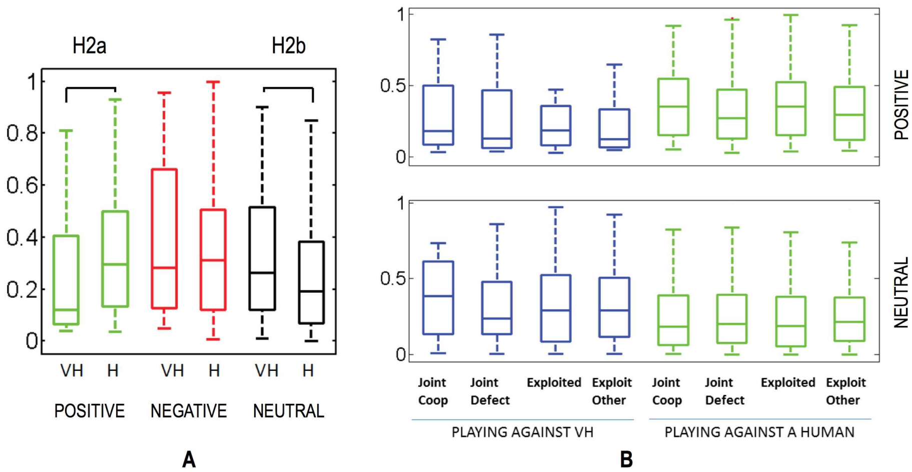

Fig. 4. Comparison of displays of expressions when playing with a VH or a human. Both overall (A) and when breaking down the game by game state (B), participants display more cooperative behaviors on average when playing with a human.

We show our first observations in Figure 4A, namely that participants display a trend of higher intensity of positivity (H2a) and less neutrality (which translates to more expressivity, H2b) when playing against another human versus playing against a VH, as we hypothesized in H2. Those trends were both significant at the 0.1 level, compared with a standard T-test.

These trends combined translate into an overall more social signal that participants communicate with their expressions to other humans while VH opponents are receiving a less social treatment. This observation still holds when breaking down by different game states as seen in Figure 4 B (bottom) for expressivity and B (top) for positive intensity. These findings agree with de Melo et al.[8] reports on participants' chosen signaled affect during the game and support our second hypothesis that people will display more cooperative expressions to human players compared with VH ones.

5 Discussion

When comparing over the same social dilemma task, we demonstrated that participants will act more cooperatively towards other humans than VHs, both in terms of game choices when choosing to betray, forgive or cooperate (as described in H1) and in terms of displaying cooperative expressions such as more joy, or less neutrality (as described in H2). It can be argued that both of those aspects of behavior form a coherent profile for the players that is more social when facing other human players than when facing VH opponents. This general observation agrees with previous findings [7] that although people treat VH like a social entity, they don't treat them equally to other humans.

The observations made in this study are locally independent of the strategy used by the opponent, however due to the iterative nature of the game the overall strategy used by an opponent should be considered as another confounding factor and be further investigated.

As discussed in Section 2 the difference in behavior shown by participants against VHs may have to do with the poorer perception of emotion expression and agency of the VH [3]. Interestingly enough, in the self-report questionnaires the participants reported less connection to a VH opponent (M=2.70, SD=1.49) versus a human opponent (M=4.46, SD=1.73), $t(112)=4.48$, $p<0.001$. This less-felt rapport could explain why participants display more neutral expressions while interacting with a VH than with a human opponent. However, considering also the communicative, coordinative role that facial expressions play, one can hypothesize that the knowledge or the expectation that the VH cannot receive these signals the same way as a person does, would lead a person to allocate less effort into that signaling channel and perhaps to cooperative behavior over all. This may be tied with the observation that when playing with the VH, participants scored significantly less than when they were playing with humans (H: M= 44:10, SD=13.24; VH: M=32.83, SD=10.86), $t(111)=3.77$, $p<0.001$. Understanding those gaps better is a topic of future work and it would help bring VH interactions closer to human-to-human ones.

6 Acknowledgements

This research was supported in part by the AFOSR [FA9550-14-1-0364] and the US Army. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

References

1. Marian Bartlett, Gwen Littlewort, Tingfan Wu, and Javier Movellan. Computer expression recognition toolbox. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–2. IEEE, 2008.
2. Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. *Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training*, volume 8108 of *Lecture Notes in Computer Science*, chapter 10, pages 116–128. Springer Berlin Heidelberg, 2013.
3. Jim Blascovich. A theoretical model of social influence for increasing the utility of collaborative virtual environments. In *Proceedings of the 4th international Conference on Collaborative Virtual Environments*, pages 25–30. ACM, 2002.
4. Joost Broekens, Maaike Harbers, Willem-Paul Brinkman, Catholijn M Jonker, Karel Van den Bosch, and John-Jules Meyer. Virtual reality negotiation training increases negotiation knowledge and skill. In *IVA*, pages 218–230. Springer.
5. Celso de Melo, P J Carnevale, S. J. Read, and Jonathan Gratch. Reading peoples minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, 106(1):73–88, 2014.
6. Celso M De Melo, Peter Carnevale, and Jonathan Gratch. The influence of emotions in embodied agents on human decision-making. In *IVA*, pages 357–370. Springer, 2010.
7. Celso M de Melo, Jonathan Gratch, and Peter J Carnevale. The effect of agency on the impact of emotion expressions on people's decision making. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 546–551. IEEE, 2013.
8. Celso M de Melo, Jonathan Gratch, and Peter J Carnevale. The importance of cognition and affect for artificially intelligent decision makers. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.
9. Jesse Fox, Sun Joo Ahn, Joris H Janssen, Leo Yeykelis, Kathryn Y Segovia, and Jeremy N Bailenson. Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *HumanComputer Interaction*, pages 1–61, 2014.
10. Sören Krach, Frank Hegel, Britta Wrede, Gerhard Sagerer, Ferdinand Binkofski, and Tilo Kircher. Can machines think? interaction and perspective taking with robots investigated via fmri. *PloS one*, 3(7):e2597, 2008.
11. Gwen Littlewort, Jacob Whitehill, Ting-Fan Wu, Nicholas Butko, Paul Ruvolo, Javier Movellan, and Marian Bartlett. The motion in emotiona cert based approach to the fera emotion challenge. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 897–902. IEEE, 2011.
12. C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000.
13. Clifford Nass, Jonathan Steuer, and Ellen R Tauber. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78. ACM, 1994.
14. Byron Reeves and Cliff Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press, New York, 1996.
15. René Riedl, Peter Mohr, Peter Kenning, Fred Davis, and Hauke Heekeren. Trusting humans and avatars: Behavioral and neural evidence. 2011.

16. Gerben A Van Kleef, Carsten KW De Dreu, and Antony SR Manstead. An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in experimental social psychology*, 42:45–96, 2010.
17. Astrid M von der Pütten, Nicole C Krämer, Jonathan Gratch, and Sin-Hwa Kang. it doesnt matter what you are! explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6):1641–1650, 2010.