Adaptive Virtual Reality based on Eye-Gaze behavior

by

Pascal Remeijsen

to obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Thursday January 19, 2023 at 10:00 AM.

Student number: 4286243

Thesis committee:
Dr. ir. Willem-Paul Brinkman
Dr. Myrthe Tielman
Prof. Dr. Marcus Specht

An electronic version of this thesis is available at https://repository.tudelft.nl/.

Preface

This thesis is a part of the Computer Science master program at the TU Delft. The journey of this thesis turned out to be longer than anticipated. During the writing of this report, an unprecedented situation developed. The Covid-19 pandemic was an event nobody saw coming, and had a significant impact on the progress of this thesis.

The original plan for my experiment was to use a lot of participants. Because this became impossible due to the strict measures put in place by the government, I had to reduce the scope of the experiment. It was changed in a way that I could do an experiment with less participants. I hope that the result is still an interesting read.

I would not have made it through without the support from friends and family. First of all, I would like to thank Willem-Paul Brinkman, my supervisor. He showed me a lot of patience and gave me good counsel. I want to thank my family for giving me the space and time to work my way through this experience. Special thanks to Mitchell, for all his support.

Pascal Remeijsen
Delft, january 2023

Abstract

Virtual cognitions are simulated inner thoughts, which are presented as a voice over. Previous research has shown the ability of virtual cognitions to increase the self-efficacy and knowledge of the users. When presenting such VR systems to users, having the VR system adapt to the user can improve their efficacy. For instance, by using eye-gaze tracking in order to adapt the VR scenario based on what the user is looking at. The measures of the ownership and plausibility of virtual cognitions were found to be important in previous research. The research described in this thesis has incorporated a social aspect into a gaze-adaptive VR system with virtual cognitions. A system was designed which puts the user in a social VR scenario where a dialogue between 3 virtual characters is shown. The user watches and listens from the perspective of one of the 3 characters. An experiment was carried out over consecutive days, where each day a new social scenario would be used. The scenario of that day would be shown twice. Once being gaze-adaptive, and once non-adaptive. After each of the 2 viewings of the VR scenario, the participants would fill in a questionnaire to measure ownership and plausibility. The results did not show a significant difference between the gaze-adaptive and non-adaptive scenarios on ownership and plausibility. The eye-gaze is determined by using special VR-goggles, with built-in cameras, capable of measuring eye-gaze.

Contents

Chapter 1

Introduction

In this chapter the context, motivation and research questions for this thesis are outlined. Next, the approach carried out in this thesis is explained.

1.1 Context

Long gone are the days in which people think that Virtual Reality (VR) systems are only good for playing games. In the last decade, there has been an enormous influx of VR systems meant for more serious goals [23]. From exposure therapies related to phobias [5], to systems for the practicing of social cognition in children with autism [7]. During the last few years, researchers have been exploring new ways to improve the efficacy of these serious VR systems. From improving the hardware, to focusing on the psychological aspects like immersion [6], or a combination of the two. One example of this, is using a special VR Heads-up-display (HUD) to determine where the user is looking in a VR environment, and using that information to improve the experience in some way [39]. This eye-tracking technology can be used in various ways. For instance in the medical field, using it to evaluate cognitive ability [34].
One example of a novel concept in VR is called Virtual cognitions [10]. Virtual cognitions are a simulation of inner thoughts, which are presented as a voice over. Users record lines of text, either dialogueue or internal monologue, in their own voice, and this audio is played back to them within an immersive VR scenario. The goal is to provide a vicarious experience as an easier point of entry for participants that find active VR training too intense.

When using virtual cognition's, the script can range from motivational statements to dialogue. This concept can be used to improve self-efficacy, provide words of encouragement and more.
In the field of VR research the psychological concepts like immersion are not always strictly defined [6]. Terms are confused or used interchangeably. One example of a term often used together with immersion is sense of presence. In the work by Slater et al. [29], Immersion is said to refer to the objective level of sensory fidelity a VR system provides. Sense of presence is said to refer to a user's subjective psychological response to a VR system. For the purposes of this thesis, the sense of presence is the most relevant concept [35].
After researching the use of virtual cognitions in negotiation training, Ding et al. investigated the effect of including eye-tracking in a VR system [12]. The gaze-adaptive VR determined the focus of attention, what the user is looking at in the VR environment, in order to influence how the virtual cognitions were presented. It was shown that the gaze-adaptive behavior can influence the sense of presence. Despite the fact that the relation between sense of presence and the outcome of VR therapy not being proven, as some studies fail to find a relation [22], there are other reasons for caring about the sense of presence. The work by Ling et al. [25], found a connection between sense of presence and self-reported anxiety.

1.2 Motivation and problem description

The success of VR systems with serious goals, depends on a variety of factors. One of them is the sense of presence that they can provide. In the scientific literature, there are a number of ways described in which the sense of presence in VR systems was investigated, for example the form in which the content is provided, like Heads-up-display compared to a projector, or the inclusion of affective content [2]. In the aforementioned work by Ding et al. [12], two derivatives of the sense of presence are used to examine the effects of the virtual cognitions. These are the plausibility of the scenario and the ownership over the virtual cognitions. One example of a question to determine the plausibility of a scenario would be: "How appropriate and natural was the voice you heard?". One example of a question to determine the ownership of a virtual cognitions would be: "Do you feel a strong connection with the voice you heard?". The result of a Bayesian analysis was

that the ownership had credibly improved over the non-adaptive condition. For plausibility, a weak improvement was found.
Whereas the work by Ding et al. involved exposure to virtual animals, this thesis will investigate whether the plausibility and ownership are also effected in a social scenario setting by implementing gaze-adaptive behavior. What differentiates the animal-exposure research and this thesis, is the presence of dialogue. Therefore, the question is whether making both the flow of the dialogue, as well as the virtual cognitions adapt based on the eye-gaze of the user will have a positive effect.
The question now becomes, how exactly can we adopt the flow of dialogue and the virtual cognitions depending on the eye-gaze of the user. Although virtual reality technologies using virtual characters and avatars have shown promising results [33], a possible point of improvement is that the avatars and characters often behave independent from the experience of the user. There is often a misalignment between the content of the VR and the focus of attention of the user. It has been suggested that this misalignment could negatively impact the sense of presence [8]. It can take the participant out of the experience and decrease the plausibility. One of the strategies of relieving these problems is the use of eye-gaze tracking. The hypothesis is that by tracking the eye-gaze of the participants, the scenario could adapt to decrease the misalignment. If the scenario changes depending on their interaction with the environment, this could give the participant the idea that they are in a more realistic scenario. However, currently there has not been much research about how to implement the gaze-adaptive behavior.

1.3 Research questions

The research questions can be formulated in the following ways:
With the goal of influencing the perceived plausibility of the scenario and ownership of the virtual cognition,

- How can one adapt the virtual cognitions in a social scenario setting depending on the user's eye-gaze?
- How can one adapt the conversational flow among the virtual characters, depending on the user's eye-gaze?

- What effect do the gaze-adaptive virtual cognitions and gaze-adaptive conversational flows have?

Based on previous research, the hypothesis is that the gaze-adaptive behavior will have a positive effect.

1.4 Approach

To answer the research questions, a number of steps were taken. First, in the next chapter, a review of previous research was done to gain insight into the problem at hand. For the first two research questions, the existing literature is consulted for possible strategies. With the findings from chapter 2, a system was designed and is presented in chapter 3, where a solution to the problem statement is laid out. The choice was made to go for a queue-based system. All recorded lines of dialogue and virtual cognitions are loaded into a queue. In the non-adaptive condition, you go through the list in sequential order. In the gaze-adaptive condition, looking at certain subjects in the VR environment will cause new snippets of dialogue or virtual cognitions to be inserted into the front of the queue, which means that those will be heard first, before returning to the rest. The implementation uses virtual reality goggles which can measure eye-gaze and transports the participant to a social setting. In chapter 4, an experiment is described, where the effects of the chosen implementation are measured. The experimental design took inspiration from studies with repeated measures. For 5 consecutive days, a comparison was made between a gaze-adaptive version of a VR social scenario, and a non-adaptive version of that same scenario. Each day, a new scenario is used. Finally, in chapter 5, conclusions are drawn, and possible future work is laid out.

Chapter 2

Related Work

In this chapter, the results of a literature study are presented. This study was done to gather domain knowledge. Specifically, the scope and requirements of the system need to become clear in this chapter.

2.1 Previous work

A good place to start, is to look at what has been done before with the main components of the system. These include the VR environment itself, eye-gaze measurements, virtual cognitions, voice-recording and more.

2.1.1 Virtual Reality Systems

VR technology has moved to the forefront of the tech-industry. More and more companies are investing in making VR cheaper, and making it available for a wider range of consumers [1]. This consumer technology is also being used for a wider range of application than ever before [18]. Hamad et al. describes a list of applications that seems to grow every year. From driving simulations, to product prototyping, to education, medical training and therapy. Similarly, in academia the adoption of VR seems to be going strong. In Pan et al. [32], a number of reasons for the adoption of VR in academia is given. First, VR gives researchers vastly more control over the experimental environment, compared to an experiment done the classical way. Because the VR environment is a software product, experiments are much easier to reproduce, as the entire environment can be shared with other researchers.

When an experiment is replicated in the real world, the method is copied, but a lot of factors like location and circumstances are very different. Another advantage is that in VR a situation can be shown that in real life would be to dangerous or impractical.
Although VR seems to have a bright future, the technology introduces new challenges. The quality of the experience seems to hinge on a number of factors. Avatar realism is one of those factors [38]. This can pertain to the appearance of an avatar, as well as to behavioral realism. It was shown that when an avatar is not sufficiently realistic in these 2 domains, it negatively impacts presence. Besides the appearance of VR avatars, seemingly minor details such as gazing behavior have been shown to affect the believably of the VR experience [15]. An avatar with gaze-aware behavior was perceived to be more socially present than an avatar that did not have such behavior. Work by Kang et al. [21], has shown that if the presence is high, this has a positive impact on factors such as feelings of Empathy with virtual avatars [40].
This research leads us to our first requirements. In order to create a sense of presence, both the content of the social scenario, as well as the delivery of the content should be relatable and realistic to the participant. This means no outlandish scenarios that people would not see in daily life. In terms of the way the content is shown, reactions should be clearly linked to actions. Because the focus of this research is not on the improvement of the virtual assets, the obvious way to link action and reaction, is to make the time between them small. Because this research deals with a social setting through dialogue, this implies that the recorded lines of dialogue should be cut up in small chunks, so that the script can quickly adapt to the gaze-behavior of the user of the system.

2.1.2 Virtual cognitions

A large number of the immersive virtual reality technologies focus on providing an experience of practice. The idea is that practicing in a virtual environment will provide an experience that is close enough to reality, so skills learned in the virtual environment can be applied in real-life situations. Besides the skills that have to be taught in some way, it is also of paramount importance that the person undergoing some intervention develops a believe that they are able to handle the problem, i.e Self-efficacy [26]. Immersive virtual reality can provide the researchers with great control over the environment [36]. However, for researchers and developers to make a scenario in virtual reality that mimics real-life in such a way that effects are transferable, seems to be a complicated task. That is partly why researchers have also searched for other avenues to pursue when trying to equip people with the needed tool to accomplish their goals. Active practise in VR is not the only strategy. One alternative is to make the experience passive. Allow the user to experience VR without having to provide any input.

When attempting to influence, for example, the self-efficacy of users, or their attitude towards the VR content, it is key that the experience is believable enough to invoke the emotions and thoughts the participant would experience normally. Providing experiences relevant to the targeted problem is one of Banduro's 4 key strategies of increasing self-efficacy [27]. This strategy is often utilized by the immersive virtual reality systems, as it provides researchers and developers with the opportunity to exercise great control over the practice environment [32]. Similarly, the environment provides a situation where there is no or less negative consequences to failing, which is a crucial part of learning. Other strategies include social persuasion, where people in the social environment can give compliments and acknowledgements to stimulate the desired behavior, as wells as psychological factors. The latter entails that the foremost factor that shapes beliefs are emotions. Influencing the emotions felt in some scenario might help shape constructive beliefs.

Finally, a strategy that seems to have been explored less, is the vicarious experience [21]. The thought that is often used in this case is: "if they can do it, so can I". This strategy is most effective when the recipients of the vicarious experience can see themselves in the person going through the actual experience. One example of when this strategy might be very useful, is that sometimes anxieties are so strong that even an active experience in virtual reality is too much to handle. When the recipients of the experience can not calmly go through the scenario, it is highly unlikely that positive and constructive beliefs will be formed.

One alternative to the active VR experience is virtual cognitions. This concept was already briefly explained in the previous chapter. The concept of virtual cognitions has been used in a few setting. The first that is discussed here is negotiation training [10]. In particular, the goal was to use virtual cognitions in an effort to improve Self-efficacy during negotiation training. As negotiation is often adversarial in nature, virtual cognition's are used to provide a less threatening alternative, while still having some of the benefits of virtual reality based treatments. The study showed a positive effect on both Self-efficacy and knowledge about negotiation. Later work by Ding et al. consisted of a larger study, which can be consulted for a more detailed analysis [9]. The general idea of what they did is as follows. They invited participants to wear a Head-mounted display, and immersed them in a virtual negotiation scenario between the character embodied by the user, and another virtual character. This experience was completely passive, and allowed the user to experience a negotiation without the pressure of participation. Throughout the experience, the user was presented with virtual cognitions. These consisted of self-motivational statements and negotiation knowledge. This larger study also found an increase in self-efficacy.

2.1.3 Eye-Tracking and gaze-adaptive behavior

In the last few years, there is a concept that has been gaining traction in the research of VR therapy system. Namely, the measurement of eye-gaze is becoming a valuable factor in experiments. There is a considerable amount of research into the use of eye-gaze in VR research. The tracking of the eye-gaze is used to make the interaction more natural, as suggested in the work by Bee et al. [3]. Similarly, incorporating the eye-gaze can improve the overall communication [16].
VR systems in general, and virtual cognitions specifically, run the risk of introducing a conflict between the cues from the environment [8]. In the work by Larsson et al. [24], it is suggested that the conflict between audio and visual cues from the environment might decrease task performance. Furthermore, when having such a conflict, it is suggested that the visual cues take precedent over the audio cues. Considering the virtual cognitions rely on that audio cue, this is an indication that striving to improve the alignment is essential. The work by Kang et al. [21]. A relation was shown between Self-identification and the sense of presence. Self-identification can take the form of self-world comparison. When there is a conflict between the cues coming from the environment, this may negatively impact this comparison. Through a negative effect on the self-identification, the sense of presence is likewise negatively affected. As mentioned before, the vicarious experience is most effective when the recipient can identify with the model. So when the model is the person themselves, the expectation would be that the effect would be at its most pronounced. But that also means that conflict are highly noticeable, as they are compared directly to ones own experience.

After the virtual cognitions were tested in the context of negotiation training, the subsequent work of Ding et al. sought to incorporate the promise of eye-gaze tracking into the virtual cognitions [12]. In this study, the context was exposure therapy. Specifically, exposure to the presence of feared animal like snakes. The authors presented the participants with a virtual encounter with a snake. During the encounter, the participants were presented with 3 types of virtual cognitions. Factual statements about the snake, self-motivational statements and instructions to encourage specific-gaze behavior, In this study, the authors chose to use the plausibility and ownership of the virtual cognitions. The plausibility illusion is the relevant component of Presence[41]. The plausibility illusion pertains to whether the participant is convinced that the VR scenario could really be happening. Specifically for the virtual cognitions, it is the believe that the audio content could be an actual cognition. That is, a snippet of internal monologue. The other measured variable ownership, as explained in the work by Gallagher et al.[14], deals with the sense of being the one who underwent an experience. Precisely because the experience is passive, the two variables will be distinct. When someone makes an active effort, these variables will overlap. But in the case of virtual cognitions, the aim is to create a sense of ownership, despite the action being involuntary. The results of the study showed that the eye-gaze adaptive virtual cognitions might be an effective way of going about improving the ownership over the virtual cognitions.
This research presents further requirements. The gaze-adaptive behaviors should fit within the story of the social scenarios.

2.2 Vision and Requirements

The goal of this research is investigating the effects of eye-gaze adaptive behavior on both the flow of dialogue and virtual cognitions in a social setting. The scope of the solution is limited to methods of adaptive behavior. Meaning, no work was done on either the hardware components of the VR set-up, nor was there development on the VR assets used in the experiment.
The first two research questions pertains to the question of how exactly can the eye-gaze adaptive behavior be implemented in a social setting. The vision for the implementation is therefore centered around how the dialogue can be changed to create an improvement in plausibility. Considering the fact that the focus of attention, what the user is looking at in the VR envi-

ronment, is rarely unchanged for long periods of time [28], this implies that the implemented system in this research needed to be able to change the flow of dialogue quite rapidly. As discussed before, this implies that the recorded voice snippets should be short.
Another requirement, is that for the non-adaptive case, the topic can not be too jarringly different from the topic of the social scenario, as a big discrepancy might hinder the comparison between the adaptive and non-adaptive case. If the difference is shocking, it will always be a false positive, when trying to determine if there is change in the dependent variables.
In order for there to be an interesting flow of dialogue, we expand the number of virtual characters. Using 3 virtual characters instead of 2 will open up more ways of gaze-adaptive behavior.
For the virtual cognitions, the strategy use in the work of Ding et al. [12], is easily translated over to the social setting of this research. The authors described how looking at the snake can trigger certain virtual cognitions. The same can be done in the social setting. Looking at characters or object in the environment can trigger certain virtual cognitions.
Finally, the virtual environment itself should fit together with the subject matter of the social scenarios. If a light heared conversation is held in a gloomy place, this will feel like a misalignment. Therefore the location should be general enough to allow for multiple types of social interaction. Because we do not want the novelty of the VR environment to be a factor, a space should be created where the participant can get used to the VR. In summary, the following requirements have been found:

- The recorded lines of dialogue should be short, to allow for rapid changes.
- The subjects of the social scenarios should not be outlandish and easily relatable.
- The gaze-adaptive behaviors should fit within the story of the scenario, or be connected to the direct virtual environment.
- The VR environment should be general enough to accommodate multiple scenarios.
- Besides the virtual space where the experiment takes place, also create a place for the participant to get familiar.

- Use 3 virtual characters.

With this list of requirement, a VR system was created with the purpose of carrying out an experiment.

Chapter 3

System Design

To answer the research questions, a Virtual Reality experiment was designed. The experimental design is explored in chapter 4. In this chapter, the design of the VR system is laid out, as well as the implementation of the system. The first section is an anecdote about the experience of the participants in this experiment. In the following section, an explanation is given for how the experience is implemented.

3.1 User experience

In order to make clear what happens during the experiment, we start with a walk-through of the experiment from the participants point of view.
On the first meeting, lines of dialogue and virtual cognitions will be recorded. After the recording is done, the partipcant will use the VR system for the first time. The room is the same as during the experiment, but the object and characters are removed.
On the 5 days of the experiment, when a user puts on the HUD, they are greeted by a room filled with 2 virtual characters and many objects. The particpants sees and watches from the perspective of a third virtual character. The particpant will not have a body. After a short delay, they will start hearing dialogue between the characters in the room. Perhaps the first one to speak is the character that the user is embodying. This means that the user hears lines of dialogue in their own voice. Next, the other characters will chime in, and have a conversation.
In the case that you are in the gaze-adaptive condition, the following might

happen. During the conversation you look away and notice a painting on the wall. As you look at the painting, one of the other characters make a remark about it, before returning to the conversation.
At certain points throughout the VR event, the user hears a virtual cognition. To their surprise, it is about the statue on the table that you were just looking at.
At the end of the conversation, the researcher informs the user that the event is over, and that they can take of the HUD.

3.2 High-level System Design

From the research questions mentioned in Chapter 1, as well as the information gathered from previous research, a set of requirements was created, as described in chapter 2. To assess the impact of gaze-sensitivity on a social scenario in VR, we need to be able to compare a eye-gaze adaptive scenario with a non-adaptive scenario, where the gazing behavior has no influence whatsoever.
The first requirement is to create a VR environment in which to conduct this comparison. One of the requirements, is the use of a space that is general enough for multiple scenarios. The setting of a conference room was chosen, because it is both general and social in nature. Besides the conference room where the experiment takes place, a waiting room is also be included. This fullfills the requirement of a space for the participant to get used to VR. The purpose of the waiting room is to allow the participants a moment to make themselves familiar with the VR setup. This opportunity was on the same day the recording of the text used in the experiment took place, which was before the real experiment began.
Next, virtual characters were needed for the social component of the experiment. The participant interacts with two virtual characters. The virtual characters have dialogue, which is recorded by individuals that were unfamiliar to the participants. The participants plays one of the three characters in the social scenario. Prior to the experiment, lines of dialogue and virtual cognitions are recorded. To run multiple experiments, a script with multiple scenarios is needed. In total, 5 social scenarios were written. Finally, the room was also filled with objects that can serve as targets for the participants to look at.
The next requirement pertains to how the gaze-adaptive behavior is ex-

pressed. What effect does looking at the objects and characters in the environment have? Likewise, the question of how the cognitions and dialogue could flow together had to be answered. For the cognitions, different versions were made for the non-adaptive case and the adaptive case. As per the requirements, the difference could not be too jarring. This meant that the non-adaptive virtual cognitions would comment and the subject matte of the scenario. This way it flows with the dialogue. For the gaze-adaptive case, virtual cognition can be about object in the room. For the dialogue, a queue-system was created to allow for the interjection of relevant snippets of dialogue in the script. In the adaptive case, looking at certain items might trigger these different dialogue snippets.

3.3 System

In this section, the implementation of the system will be discussed. Including the technologies, content of the VR experience and the consideration for the gaze-sensitive behavior.

3.3.1 Technology

To carry out the designed experiment, a set-up consisting of various hardware and software components had to be chosen. The hardware consists of:

- Virtual Reality Head-mounted Display (HMD) [Fove Zero].
- In-ear microphone and headphone.
- Laptop [Alienware Gaming laptop].

The software consists of:

- Unity (Game Development Software)
- Audacity (Audio tool)
- UMA Character packages + Realistic movement packages

3.3.2 VR Assets

A number of VR assets were utilized, ranging from character models to everyday object. A lot of the assets were re-used from previous research, including most objects [15]. Most were obtained from the free VR assets database called Archive3d[1]. These assets were combined to form an office setting. An office setting would offer the most versatility, as the environment is dictated by the social setting. A number of different scenarios could plausibly be presented. The setting is a room with a door and a window looking out at other buildings. The room was inhabited by a table, 3 chairs, for 3 characters and an assortment of object like picture frames, a painting, a statue, a coat rack and more. These items will present opportunities for gaze-behavior. In figure 3.1 through 3.3, you will find an image of the VR environment.

[1]https://archive3d.net/?tag=paper

Figure 3.1: View of the front of the VR environment

Figure 3.2: View to the side of the VR environment

Figure 3.3: View from the top of the VR environment

The waiting room consists of the same physical space, but notably without the characters. Because natural gazing behavior is desired, the rooms must not be too similar, as not to influence the exploration behavior in the main room. Thus, the objects in the room will also be changed.

3.3.3 Characters

There were no special requirements to the appearance of the virtual character models themselves. Previous research has shown that realistic head-movement is a factor in the plausibility of the model[19]. The UMA-framework[2], together with a "Realistic Eye Movements" Unity Package [cite correctly] was chosen to enable realistic model behavior. The choice was made to include 2

[2]http://www.project-uma.com/the-uma-model

VR characters in each scenario, to give more options for the focus of attention. This allowed for more possibilities in how to implement the adaptions, and offers more ways to direct the dialogue.
The participant can view the social interaction from a seating-height, but does not actually physically embody a character, the entire experience is passive. The participant is able to turn 360 degrees to look around the environment.
The virtual character across from the participant will be 1 male and 1 female. The choice for this set-up was made because distinguishing between two different sex voices is much easier. The third voice, obviously being the participants own voice.

3.3.4 Dialogue script

The script is the collection of all lines of text that have to be recorded for the voice-over of the VR experience. The script contains different types of text, including virtual cognitions, dialogue and interjections.
In both the gaze-adaptive and non-adaptive conditions you will hear virtual cognitions. For the virtual cognitions, separate versions for both the non-adaptive and adaptive cases were written. In the adaptive case, the topic of the virtual cognition is determined by which object the user is currently looking at, or looked at last. This type of text takes the form of either comments about an object in the environment, or the virtual characters across from the participant. The key issue for the non-adaptive case, is that the topic can not be too jarringly different from the topic of the social scenario, as a big discrepancy might hinder the comparison between the adaptive and non-adaptive case. When the scenario pertains to a social scenario, and the cognition mentions polar bears, this might feel as an unnatural interjection, which would surely influence the result. Therefore, the non-adaptive virtual cognitions were chosen as to fit into the subject of the scenario.As shown in table 3.1, the cognitions for the adaptive case revolves around virtual objects. In the non-adaptive case, the cognitions are related to the topic of discussion in the social scenario. Hearing a thought about the topic of discussion, would certainly not break the plausibility of the scenario.

The dialogue script are the main lines of text that the participants will hear. In the case of the non-adaptive condition, the participant will hear this list of lines in sequential order, without interjections. There are scripts for 5

Adaptive	Non-adaptive
"What is the deal with that statue. Shouldn't they be displaying this in a case somewhere"	"This conversation is giving me hope. I am certain they can still turn things around."
"I don't like this painting. They should have displayed one with nature in it."	"I was afraid they were going to say that. This is going to be difficult"

Table 3.1: Examples of non-adaptive and adaptive cognitions.

scenarios. All scenarios were designed to be a social setting, which allowed for dialogue between 3 characters. The scenarios chosen were: A discussion about a promotion, buying a house, adopting a dog, a discussion about an exam and finally a talk with a medical professional.

Let's run through an example. In figure 3.2, you will find the start of a script pertaining to a scenario about buying a house. This is a social scenario, where a couple talks to a Realtor. In this script, you can see the dialogue, as well as three virtual cognitions, indicated by the italics. These cognitions pertain to the topic at hand which means that this example is a non-adaptive case. In the next section, it is explained what it would look like if it was an adaptive case. What is not displayed here are the dialogue interjections. These will be explained in the subsequent section.

Speaker	**Line**
User	Hello and welcome. I am glade you guys are here.
Character 2	We are pleasantly surprised that you could see us so quickly. You guys must be busy.
User	That is true. We are very busy. But we had an opening for you. How can I help you?
Character 1	We are in a hurry to find a new home.

Table 3.2: Script for a scenario about buying a house.

After an introduction of the topic, now comes the first virtual cognition that is used. This should be a cognition that is on topic in the case of the non-adaptive condition. See table 3.3.

Virtual Cognition	**Line**
User	*I hope they understand that it is difficult to find a home under these market conditions. They should be happy if I can find anything.*
User	*I was afraid that was the case. This is going to be very difficult.*
User	*I am glad they don't have outlandish demands. I think I will be able to find something.*

Table 3.3: Examples of virtual cognitions that are on-topic.

The cognitions shown in 3.3 are played on predetermined positions in the non-adaptive case. The rest of the dialog is shown in 3.4.

Speaker	Line
User	Is there a specific reason for that?
Character 1	We have to leave our current residences quite soon.
Character 2	We have been wanting to move for a while, but now we really have to move.
User	I understand. That is a tough situation. The problem is that demand is very high at the moment. What are you looking for.
Character 1	We are looking for an apartment with at least 3 bedrooms and a large kitchen.
User	Any other specific criteria?
Character 1	Our budget is around 300,000.
User	Are there any other requirements that you didn't have in the old house?
Character 1	The location is not ideal. We would like to live close to a school.
Character 2	A balcony would also be nice.
User	I think we are ready to start our search. You can't demand too much in this market.
Character 1	I hope we can visit a residence soon.
Character 2	We are willing to be flexible.
User	That is good to hear. I will start the search, and you guys will hear from me soon.

Table 3.4: Script for a scenario about buying a house.

3.3.5 Dialogue interjections

In the adaptive case, the flow of dialogue is altered by the gaze behavior of the participant. This was accomplished by implementing a queue-system. This means that at the start of the scenario, all lines of text are queued up. Then in the case of an event, which means looking at a certain object, a line of text pertaining to that object can be put at the front of the queue, and thus

heard immediately. In the non-adaptive case, a predetermined script is laid out. This script will not change during the duration of the VR experience. This includes the content of the lines of dialogue, as well as the turn-taking between the participant.
In short, the queue-system allows new lines of text to be interjected into the script. As seen in table3.5, these interjection will belong to two categories. Either returning the attention back to the dialogue, or commenting on the object that the user looked at.

Joint attention	Return attention to dialogue
"I hope that lamp will be replaced soon"	"Excuse me, are you listening?"
"I thought that picture was weird too"	"Sorry, are you talking to me?"

Table 3.5: Examples of Interjections

3.4 Implementation details

In this section an explanation is given for how the gaze-adaptive behavior is implemented in the system. First, lets look at the dialogue interjections. The system is implemented in Unity. In Unity, each object in the environment can be made to run snippets of code, called "scripts". A script can determine the behavior of the object, and enact changes in the environment. A script called "Soundmanager" is attached to the speaker-object, which is responsible for producing all the audio that the user hears, and contains all lines of text that belong to the dialogue in a scenario.
A "collision" script is attached to various objects in the environment. This script causes a message to be send to the "Soundmanager" script, whenever the measured eye-gaze collides with an object. The way this works, is that a standard package for the particular VR goggles used in this experiment, draws lines from the "player" object out toward the focus of attention. If the beam collides then some action is triggered. This trigger lets the "Soundmanager" know to insert the line of text associated with that object into the front of the queue. This means that that line will be played first.
There are a number of considerations in this set-up. It is possible that the start of a VR experience entails the participant frantically looking around.

It is undesirable for all of the interjection to happen before the dialogue gets going. To mitigate this, it is enforced that there can be no two interjections following each other. This ensure that the interjections are dispersed among the lines of dialogue.
For the adaptive cognitions, a pointer to the most recently viewed object is kept, which in turn decides which cognitions is heard at the fixed places in which a cognition is played. When a cognitions turn is up, it will look at the last viewed object, and play the line of text associated with it. In the non-adaptive case, the cognitions are part of the fixed script, and there is no influence on this from the user.

Chapter 4

Evaluation

After the development of the VR system, an experiment was carried out, to answer the research questions. In the first sections of this chapter, the methods used for the experiment will be explained. This includes the experimental design, how the participants were selected, which measures were used and what kind of analysis was done. Following the methods section is the result section, where the outcomes of the experiment are explored.

As described in the previous chapter, the questions we wanted to answer concerned gaze-awareness. Mainly, whether incorporating the information of where the participant is looking in the scenario, can influence two particular independent variables. The hypotheses that were tested are as follows:

- H1: Adapting the flow of dialogue depending on where the user is looking in the virtual scenario, has a positive effect on the plausibility of the scenario compared to a non-adaptive dialogue flow.
- H2: Adapting the virtual cognitions depending on where the user is looking in the virtual scenario, has a positive effect on the ownership of the cognition compared to non-adaptive virtual cognitions.

When the design of the experiment was made, the global COVID-19 pandemic was a complicating factor. These circumstances complicated both the recruitment of participants, as well as the general organisation of an experiment with human subjects. We chose to conduct an experiment with a small number of participants, to minimize the external complications. Particularly, participants were found who already live together, which means that the experiment does not introduce additional risk. To compensate for the small number of participants, an experiment with repeated measures was required. For the design of this experiment, inspiration was taken from both Single-case studies with repeated measures[37], as well as Micro-RCT trials[4]. The details are presented in the following sections. For this experiment, the TU Delft Ethics Committee approved the human research ethics application with ID number 1532. The study was pre-registered on the Open Science Framework (OSF)[1]. The data collected through the experiment can be found on the 4TU Center for Research Data.

4.1 Methods

In this section, the design principles, as well as the practical details of the experiment will be explained. Lastly, the analysis plan is discussed.

4.1.1 Experimental Design

To examine if we should accept or reject the aforementioned hypotheses, we needed to compare the gaze-adaptive scenarios to baseline scenarios which are not gaze-adaptive. This means that the independent variable is a binary categorical variable, indicating the presence of gaze-adaptive behavior. This results in the conceptual design as depicted in figure 4.1.

[1]https://osf.io/p5cf7/

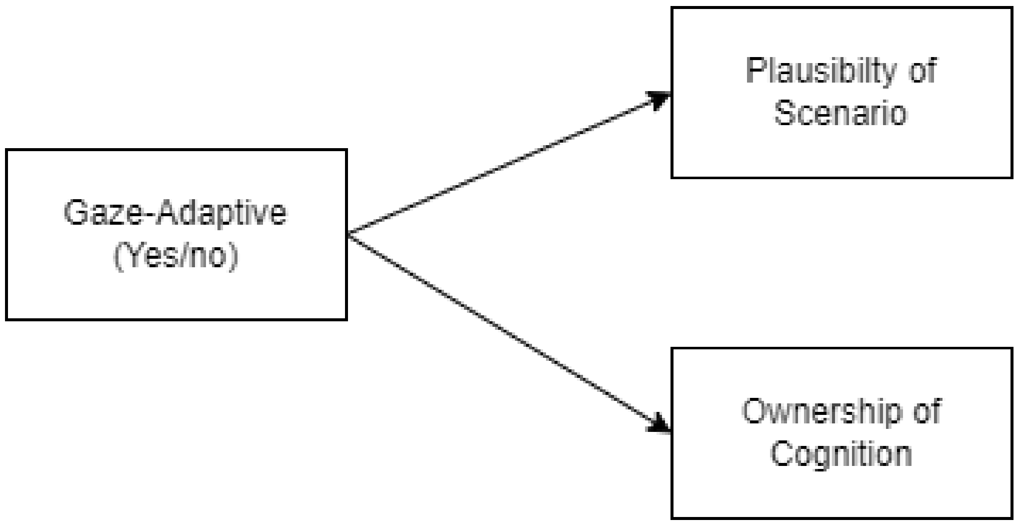

Figure 4.1: Conceptual Design.

This conceptual design requires the ability to display two versions of a VR scenario. One with the gaze-adaptive behavior and one without. Both versions have the same visual presentation, meaning the room and the objects contained within, are entirely the same in both versions. Only the auditory content of the VR scenario is determined by the independent variable. In the Gaze-adaptive case, the system presents a VR scenario that adapts to where the user is looking. This is done through both the virtual cognitions, as well as the dialogue interjections. For the non-adaptive case, the audio is completely predetermined. This means that there are no dialogue interjections at all.
Taking inspiration from Single-case studies with repeated measures[37], lead us to the decision to repeat the experiment for multiple consecutive days. With classical experiments, researchers look for systemic patterns across participants. However, with Single-case studies with repeated measure, researchers look at systemic patterns across events. In this case, an event, is one experience with the VR Systems. Because each scenario is done twice, once in an adaptive version, and once a non-adaptive, this means that there are an equal number of adaptive events, as non adaptive event. However, this set-up is robust against errors where the data of events get lost, as it is not required that there are the same number of events for each condition. So is some get lost the analysis can still be carried out. This is a viable alternative

to use in case of a low number of participants, due to the health and safety constraints at the time of the experiment.
Computer generated randomization was used to exclude confounding effects. For instance, if the non-adaptive case is always first, it is hard to distinguish whether there is a difference due to the independent variable, or perhaps there is something like a learning effect. The randomization is implemented in two ways. Both on the level of scenario topics, as well as the order of the two conditions per day. On each day of the experiment, a scenario is provided, once in an gaze-adaptive version and once in a non-adaptive version. Which of the two versions is depicted first is decided at random, with equal probability for both possibilities. Additionally, the order of the scenario topics is randomly generated on the first day of the experiment. For the following days, the initially generated schedule was used. These principles result in the following experimental design.

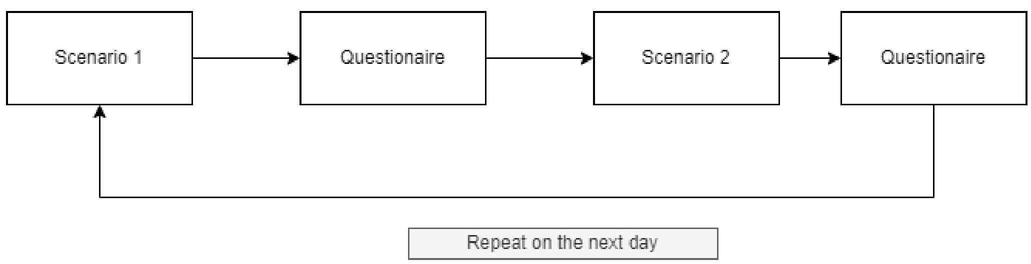

Figure 4.2: Experimental Design.

With the experimental design as depicted in Figure 4.2, there are a number of details that remain. The number of days that the experiment would run had to be decided. Together with the number of participant, this would determine the number of VR events, and thus the sample size of the experiment.
A Monte-Carlo power simulation was used to determine the number of days required to get an acceptable statistical power level. For the analysis, a Multi-Level model with Random intercepts was used. The random intercepts are the participants of the experiment. The fixed effect is the condition, which is the binary variable that indicates if the event was adaptive or non-adaptive. The justification being that the different participants likely have their own sensibilities when scoring, and are assumed to be consistent within the level of the participant. The predicted variables are on a likert-scale (-3 to 3).

The categorical predictor variable is the inclusion of gaze-sensitivity.

The Monte-Carlo simulation was run with the standard significance level of 0.05. A big problem arose during the preparation for the experiment. I should have looked at the effect size and standard deviations that were found in earlier research. Namely, during the experiment with virtual cognitions which preceded and inspired this research, as described in Ding et al.[12]. They found a mean difference for the ownership of 0.55 and 0.080 for the plausibility. For the Monte-Carlo simulation of this research a much larger mean difference between conditions, closer to 1 was used. The second mistake was to use the same parameters for ownership and plausibility. Likewise, the standard deviations was too large.
However, with the chosen parameter, the number of days was increased until a power level of 0.8 was reached. The resulting number of days that met the requirements was 5. This means 5 days of two events per day, for each participant. Resulting in 30 total VR events.
Due to the mistaken parameters for the Monte-Carlo simulation, the experiment ended up being under-powered.

4.1.2 Measures

The measured variables are the plausibility of the scenario and the ownership of the virtual cognitions. Both variables were measured by questionnaire. The questionnaire from the paper by Ding et al.[12] was used. This questionnaire consists of two sections. There are 5 statements for the plausibility and 3 statements for the ownership that the users are asked to score. The section of the questionnaire relating to the plausibility was inspired by a questionnaire used by Millevill-Pennel and Charron[30]. The section on ownership was a combination of two earlier questionnaires[20, 31]. The participants are asked to rate the statements on a Likert-scale (-3 to 3), ranging from 'Stronlgy Disagree' to 'Strongly Agree'. After the selection of the questionnaire, it had to be translated into the native language of the participant, which is dutch. A collection of four peers, who are native dutch speakers as well as fellow students, reviewed the translation and judged the accuracy of the translation to be sufficient. Meaning that the spirit of the statements was captured in the translation. The questionnaire was filled out online, using Qualtrics[2]. Qualtrics is an online platform which, among other services, provides the ability to host surveys. A link to the questionnaire was provided to the participant after every VR experience, and they were asked to fill out the questionnaire. Besides the measurement, the collected data included the age, gender and previous VR experience of the participants.

4.1.3 Procedure

The experiment arranged for the principal researcher to meet with the participants on 6 different days. The first meeting with the participants was the recording of the dialogue and virtual cognitions. The participants were asked to record a small collection of spoken lines in their native language (dutch). With the lines of the recorded virtual cognitions, the participant were asked to adjust the sound parameters to match their own perception of the inner voice, using provided software (Audacity). As explained in the work by Ding et al. [11], adjusting the sound file to sound more like how the user would hear their own inner voice, has been shown to have a positive effect on the ownership of the cognition. This meeting was also used to familiarize the participant with the VR set-up. On that same day, an introductory VR experience was given, where the participants were provided with an experience with the Head-mounted display. No auditory input was given during this introduction. To exclude the factor of novelty, the room was not entirely the same as during the experiment. There were no virtual characters, and the objects inhabiting the room were different. After the participants understood what it would look like, they were asked if the set-up made them nauseous or anxious, as some VR research has shown to be a possibility. After answering in the negative, the first session was concluded.
After a waiting period of a week, the core experiment was carried out over 5 consecutive days. Each day entailed two VR events, a scenario that was shown once with gaze-adaptive enabled and once the non-gaze-adaptive version. Which version is shown first on that day was randomly determined by a computer generated coin-flip. Both the participant and the researcher did not know which would be first prior to the start of the VR scenario. Each round was followed by a questionnaire. After the questionnaire was filled in for the second time, the session would be concluded. The participants were asked to not discuss the experience with others.

[2] Qualtrics.com

The manipulated variable is the inclusion of the gaze-sensitivity. This single binary categorical variable was expressed through changes in both the flow and content of the dialogue, as well as through adapted virtual cognitions. To summarize, in the non-gaze-sensitive case, the dialogue was a predetermined and fixed order of lines of dialogue. This includes predetermined virtual cognitions at the predetermined intervals. The gaze-sensitive case, provides the ability for lines to be interjected in the dialogue, according to where the participant is looking. The virtual cognitions will be played depending on where the participant is looking.

4.1.4 Participants

There were 3 individuals selected for the experiment, the youngest being 24, the oldest being 60. Individuals were selected that live under the same roof, as this fit best with the health and safety constraints that were in place at the time of the experiment. All 3 had the same level of experience, which was none. There were two female and one male participant. The participants were not selected with particular constraints in mind. There was no payment to the subjects.

4.1.5 Data preparation and analysis

For the statistical model, a Linear mixed-effects model was chosen. Because the participants are likely to have distinct interpretations of the experience, the participant was chosen as a Random Intercept. The ownership of the virtual cognitions and the plausibility of the scenario were the dependent variables. Each dependent variable had its own multi-level model. The fixed effect is the condition, which is the binary variable that indicates if the event was adaptive or non-adaptive.
The dependent variables are scores on a Likert-Scale. The manipulated variable is a categorical field, indicating the presence of the gaze-sensitivity. The means of the items on the questionnaire were used as the index variable.
Cronbach's Alpha was chosen to investigate the the internal reliability of the questionnaire items. The questionnaire, measuring both the feeling of ownership of the virtual cognition, as well as the plausibility of the scenario, are checked on their internal reliability. The sections of the questionnaire comprise 5 and 3 questions for the plausibility and the ownership respectively.

The Cronbach's Alpha scores can be found in table 4.1.5. The score for the ownership is considered "good"[17]. The plausibility met the minimum requirement of 0.6. The internal reliability was sufficient for the inclusion of all collected measurements.

Section	Lower CI	Alpha	Upper CI
Ownership	0.79	0.86	0.94
Plausibility	0.37	0.6	0.83

Besides the answering of the research question, additional exploration of the data was carried out. First, plots were made for each participant, where the difference between the conditions was shown for each day. This was done to gain insight in the progression of the experiment. Secondly, the same multi-level model was used to examine whether using the session-variable as the fixed-effect (which indicates if an event is the first or second of the day) performs better as predictor. This is with the same random intercepts. It is possible that a cognitive bias knows as 'anchoring'[13] has an effect on the result. This effect entails that the first event sets a base-line for the second. This phenomenon is often found in everyday life.
The R code for the analysis of the results can be found in the TU Delft repository[3].

4.2 Results

In this section, the results of the experiment are shown. There is a discussion of the results in relation to the hypotheses. Followed by an exploratory analysis, to gain further insight.

4.2.1 Hypothesis testing

Before the multi-level model was fitted, the differences in condition for each day was plotted. Considering the participants are the random intercepts, the difference per day will also be viewed per participant. This means that there are three lines, for the three participants, and each day represents a measurement independent of the other days. In figure 4.3, the differences

[3]https://repository.tudelft.nl/

in the levels of perceived plausibility of the scenario are shown. A positive difference indicates that the gaze-adaptive scenario was perceived to be of a higher level of plausibility than the non-adaptive scenario. With a negative difference, the adaptive case was scored lower. In figure 4.4 , the difference in the levels of perceived ownership of the virtual cognitions can be seen. A positive difference indicates that the gaze-adaptive scenario was perceived to be of a higher level of ownership than the non-adaptive scenario.
Starting with the plots of the differences between conditions. In figure 4.3, we can see that for all participants, the second day was a clear increase in the difference between the conditions. However, after the second day it plateaus. It is possible that the first VR experience with gaze-adaptivity was a novel experience. Then the second day the participants learned to tell the difference. In the case of the ownership, as seen in figure 4.4, there is less of a trend. It is possible that ownership depends more on the content on the scenario, meaning the story that is told during the VR event.

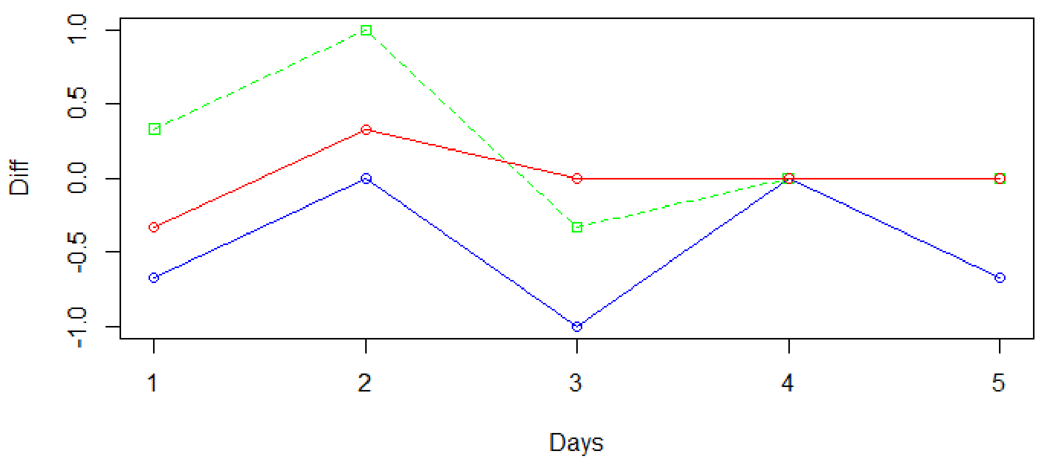

Figure 4.3: Plot of the difference in plausibility between conditions per participant.
Participant 1: Blue   Participant 2: Green
Participant 3: Red

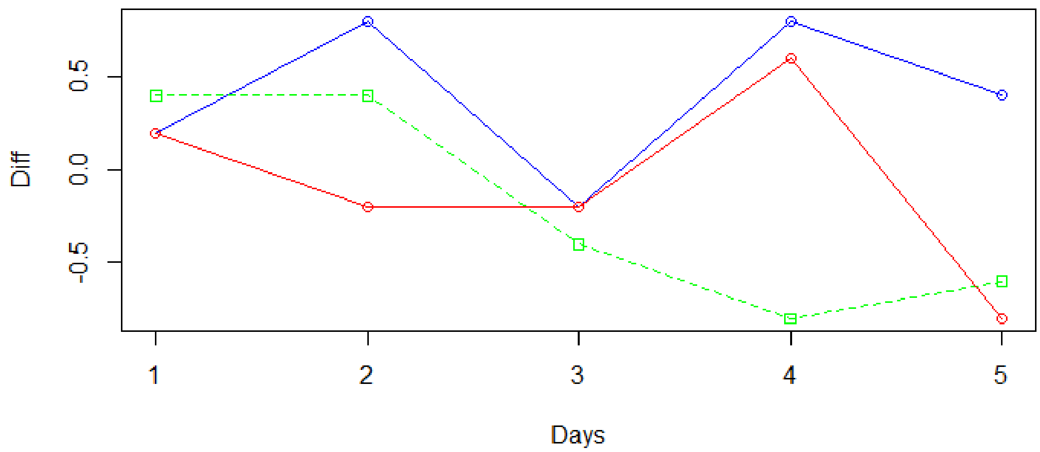

Figure 4.4: Plot of the difference in ownership between conditions per participant.
Participant 1: Blue   Participant 2: Green
Participant 3: Red

H1: Adapting the flow of dialogue depending on where the user is looking in the virtual scenario, has a positive effect on the plausibility of the scenario compared to a non-adaptive dialogue flow.

In figure 4.1, a summary of the model is given. A very small negative effect was found, which is not statistically significant.

	Estimate	Std. Error	t	p
Intercept	1.511	0.244	6.182	0
Condition	-0.089	0.146	-0.61	0.547

Table 4.1: Multi-level model of Plausibility

H2: Adapting the virtual cognitions depending on where the user is looking in the virtual scenario, has a positive effect on the ownership of the cognition compared to non-adaptive virtual cognitions.

In figure 4.2, a summary of the model is given. A very small effect was found, which is not statistically significant.

	Estimate	Std. Error	t	p
Intercept	1.84	0.281	6.55	0
Condition	0.04	0.144	0.277	0.784

Table 4.2: Multi-level model of Ownership

4.2.2 Exploratory analysis

For the exploratory analysis, we will look at whether the session variable, which indicates which VR event is first and which is second on a given day, is a better predictor. The same type of model was used as in the section about hypothesis testing, with ownership and plausibility being the dependent variables, and the Age variable being the random intercept. The predictor variable is now the Session variable. In figure 4.4, a summary of the ownership model is given. A small effect was found, which was not significant. In figure 4.3, a summary of the ownership model is given. A very small effect was found, which was not significant.

	Estimate	Std. Error	t	p
Intercept	1.444	0.245	5.906	0
Condition	0.044	0.147	0.303	0.764

Table 4.3: Multi-level model of Plausibility, with Session as predictor.

	Estimate	Std. Error	t	p
Intercept	1.773	0.280	6.323	0
Condition	0.173	0.140	1.235	0.228

Table 4.4: Multi-level model of Ownership, with Session as predictor

4.3 Results discussion

In this section, we will discuss the results that were found.

4.3.1 Answering the Research questions

H1: Adapting the flow of dialogue depending on where the user is looking in the virtual scenario, has a positive effect on the plausibility of the scenario compared to a non-adaptive dialogue flow.

The results do not support the hypothesis. The effect size is even very slightly negative, but so close to zero that no conclusion can be drawn from the sign of the effect.
Due to the under-powered experiment, resulting from a flawed Monte-Carlo simulation, we cannot conclude that this particular implementation of gaze-adaptive behavior is non-viable. It is entirely possible that a larger sample size and a properly powered study would produce different result. However, this experiment does not produce a significant effect on the plausibility of the scenario compared to a non-adaptive dialogue flow. This does not say anything about other implementations of gaze-adaptive behavior.

H2: Adapting the virtual cognitions depending on where the user is looking in the virtual scenario, has a positive effect on the ownership of the cognition compared to non-adaptive virtual cognitions.

The results do not support the hypothesis. The effect size is even very slightly negative, but so close to zero that no conclusion can be drawn from the sign of the effect. The same Monte-Carlo simulation issue is applicable here. This experiment is unable to produce a significant effect on the plausibility of the scenario compared to a non-adaptive dialogue flow. This does not say anything about other implementation of gaze-adaptive behavior.

Exploratory Analysis

The models where the Session variable (First or second VR event of the day) is the predictor are likewise not significant. Just like with the hypothesis testing, the exploratory analysis cannot conclusively say whether the Session variable is a good or bad predictor. A higher powered study might have produced different results.

4.3.2 Potential reasons for the results

In the previous sections, I have explained the issue with the Monte-Carlo simulation. But there are other potential reasons.

The first one is that the proposed implementation of adaptive-behavior is insufficient. A more radical difference between the conditions was needed to evoke a reaction from the participants.

Another issue might be the combination of the dialogue flow and the virtual cognitons. If the study design had been different, where the dialogue and cognitions where tested separately, we might have been able to gain more insight.

Finally, perhaps the questionnaire was not an appropriate tool to asses the dialog flow. In earlier research, this questionnaire was used to assess only virtual cognitions. More validation is needed.

4.3.3 Limitations

There are a number of limitations pertaining to the experiment as it was carried out. The first limitation is the possibility of the learning effect over the 5 days of the experiment. As described in the section about the experimental design, each day of the experiment, a different scenario was presented to the participant. However, the results seem to indicate that this was not sufficient to prevent participants from learning to tell the conditions apart. This could lead to a number of problems. For instance, the phenomenon where the participants start to answer the questionnaire in a way that they think will please the researcher.

In addition, it is probable that there should have been more repeated measures. As mentioned in the previous section, the study was under-powered, due to an overestimated effect size.

A complicating factor that arose during the experiment was that of the differences in ability to voice act between participants. Due to the small number of participants, personal issues have a larger effect on the results than an experiment with a larger population of participants. Some participants were unable to record the lines in a natural way, which might have effected the outcome of the ownership measurements. Additionally, the unpleasant feeling of listening to ones own voice[42], presents an effect or the earlier scenarios. The participants reported that the first time was most unpleasant, with each day being less so. In future application of the virtual cognition method, it

might be informative to ask about conditions like dyslexia, which might hinder the ability to record the dialogue in a natural way.
Both the gaze-adaptive and non-adaptive scenarios have the same amount of dialogue and virtual cognitions. However, because the gaze-adaptive scenarios also have interjections of lines in the dialogue, this case might feel slightly longer to the participant. In the ideal case, all other variables are kept the same except for the independent variable. However, in this implementation, the length of the scenarios were not precisely equal.

Chapter 5

Discussion and conclusion

In this section, the findings of this thesis will be discussed. Additionally, the limitations of these finding, as well as the derived contributions and future plans will be presented.

5.1 Findings

The research question that this report set out to answer where as follows:

- How can one adapt the virtual cognitions in a social scenario setting depending on the user's eye-gaze?
- How can one adapt the conversational flow with the negotiation agents, depending on the user's eye-gaze?
- What effect do the eye-gaze adaptive virtual cognitions and adaptive conversational flows have?

The answer to the first question is that one could use the assets in the virtual environment. This can include objects or virtual characters. the virtual cognitions could be triggered by looking at the assets. Although this seemed to be a promising option in a non-verbal setting, such as exposure therapy with animals, in a social setting this does not appear to be sufficient.
The answer to the second question, was to enable rapid changes in the dialogue flow. Adding interjections into the script of the social interaction, by either the virtual character that the user is embodying, or the other virtual characters in the environment, is one way to do that. The topics of these

interjections can be determined by which targets in the environment the user is looking at. These can again be objects or characters. These interjections can happen throughout the dialogue. Although the strategy is successful at quickly changing the flow of dialogue, on it's own, it was not capable of producing a significant effect.
Finally what effect do these implemented ways of gaze-adaptive behavior have on the plausibility of the scenario and the ownership over the cognition? The results of the experiment described in this report are not significant.

5.2 Contributions

The first contribution this research has made is in providing an overview of the concept of virtual cognitions. As this is a quite recently developed direction of research, there were no sources collecting the progress of the concept. This report has highlighted the contributions of those findings.
The second contribution is the evaluation of an implementation of eye-gaze adaptive behavior. The solution presented in this research is to connect the virtual cognitions and dialogue interjection to assets in the virtual environment. This can include objects or virtual characters. The virtual cognitions were to be played at set intervals throughout the dialogue. The item that the user is currently looking at will dictate which prerecorded cognition is played. Through an experiment it was discovered that this is most likely not enough. Even with a larger sample size, and thus a more appropriately powered experiment, this is unlikely to produce a larger effect.

5.3 Limitations

The ability to generalize the results of the experiment described in this report might be hindered by the choice of experimental design. The strategy of repeated measures was used to mitigate the health and safety constraints at the time of the experiment. However, due to an overestimation of the effect size, the study became under-powered, which means the result are not conclusive.
Additionally, The concept of looking at the patterns between events, rather than between the participant, only works when it can be assured that the

repeated measures avoid problems such as the learning effect, as well as prevent that repetitiveness influences the feedback of the participants. If the repeated actions by the participants are too repetitive, this might have an effect on the outcome.
Similarly, the number of scenarios took away from the depth and detail of each scenario. The result could have been better if the scenarios were longer and more engaging.
Another limitation of this research is that only one possible implementation was examined. This result does not hint at where the solutions may lie.

5.4 Future Work

Future work could be built around other ways of implementing the adaptability. This could mean moving away from the passive VR experience, and using active user input. It has been suggested that active VR is less prone to the users growing ambivalent towards the experience [36]. By being more engaging, the immersion and sense of presence could be improved.
One issue that should be worked on, is making the voice acting easier. This was reported to be a problem by the participants, who were not used to voice acting. Similarly, requiring participants to read out long pieces of text, excludes participants with Dyslexia. One solution for recording anxiety and other issues with recording speech maybe Deepfake technology [43]. Using the recording of natural speech, enables Deep fake models to produce a spoken form of any provided written text. This would alleviate the issues with recording.
Finally, one interesting angle of research for the virtual cognition research , is the inclusion of state-of-the-art VR assets. As shown in the Related works chapter, the previous virtual cognition's research has not focused on the visual aspect. It would be interesting to find out what the relation is between the visual fidelity and the gaze-adaptive behavior of the audio cues. Perhaps animations could accompany the audio reaction coming from virtual characters.

Bibliography

[1] Mehrasa Alizadeh. Virtual reality in the language classroom: Theory and practice. *Call-Ej*, 20(3):21–30, 2019.

[2] Rosa María Baños, Cristina Botella, Mariano Alcañiz, Víctor Liaño, Belén Guerrero, and Beatriz Rey. Immersion and emotion: their impact on the sense of presence. *Cyberpsychology & behavior*, 7(6):734–741, 2004.

[3] Nikolaus Bee, Elisabeth André, and Susanne Tober. Breaking the ice in human-agent communication: Eye-gaze based initiation of contact with an embodied conversational agent. In *International Workshop on Intelligent Virtual Agents*, pages 229–242. Springer, 2009.

[4] Tessa Beinema, Harm op den Akker, Marian Hurmuz, Stephanie Jansen-Kosterink, and Hermie Hermens. Automatic topic selection for long-term interaction with embodied conversational agents in health coaching: A micro-randomized trial. *Internet interventions*, 27:100502, 2022.

[5] Cristina Botella, Javier Fernández-Álvarez, Verónica Guillén, Azucena García-Palacios, and Rosa Baños. Recent progress in virtual reality exposure therapy for phobias: a systematic review. *Current psychiatry reports*, 19(7):1–13, 2017.

[6] Doug A Bowman and Ryan P McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7):36–43, 2007.

[7] Nyaz Didehbani, Tandra Allen, Michelle Kandalaft, Daniel Krawczyk, and Sandra Chapman. Virtual reality social cognition training for children with high functioning autism. *Computers in human behavior*, 62:703–711, 2016.

[8] Ding DING, Willem-Paul BRINKMAN, and Mark A NEERINCX. Virtual reality exposure with eye-gaze adaptive virtual cognitions. *Annual Review of Cybertherapy And Telemedicine 2019*, page 145, 2019.

[9] Ding Ding, Willem-Paul Brinkman, and Mark A Neerincx. Simulated thoughts in virtual reality for negotiation training enhance self-efficacy and knowledge. *International Journal of Human-Computer Studies*, 139:102400, 2020.

[10] Ding Ding, Franziska Burger, Willem-Paul Brinkman, and Mark A Neerincx. Virtual reality negotiation training system with virtual cognitions. In *International Conference on Intelligent Virtual Agents*, pages 119–128. Springer, 2017.

[11] Ding Ding, Mark A Neerincx, and Willem-Paul Brinkman. Simulating the inner voice: A study of sound parameters. *Annual Review of Cybertherapy And Telemedicine 2018*, page 166, 2018.

[12] Ding Ding, Mark A Neerincx, and Willem-Paul Brinkman. The effect of an adaptive simulated inner voice on user's eye-gaze behaviour, ownership perception and plausibility judgement in virtual reality. *Interacting with Computers*, 32(5-6):510–523, 2020.

[13] Adrian Furnham and Hua Chu Boo. A literature review of the anchoring effect. *The journal of socio-economics*, 40(1):35–42, 2011.

[14] Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21, 2000.

[15] Zilla Garama. The role of eye gaze in virtual characters on their perceived believability, social presence and the application of theory of mind. 2019.

[16] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 309–316. ACM, 2001.

[17] Joseph A Gliem and Rosemary R Gliem. Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales.

Midwest Research-to-Practice Conference in Adult, Continuing, and Community . . . , 2003.

[18] Ayah Hamad and Bochen Jia. How virtual reality technology has changed our lives: An overview of the current and potential applications and limitations. *International Journal of Environmental Research and Public Health*, 19(18):11278, 2022.

[19] Laurent Itti, Nitin Dhavale, and Frederic Pighin. Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation VI*, volume 5200, pages 64–78. SPIE, 2003.

[20] Andreas Kalckert and H Henrik Ehrsson. Moving a rubber hand that feels like your own: a dissociation of ownership and agency. *Frontiers in human neuroscience*, 6:40, 2012.

[21] Ni Kang, Ding Ding, M Birna Van Riemsdijk, Nexhmedin Morina, Mark A Neerincx, and Willem-Paul Brinkman. Self-identification with a virtual experience and its moderating effect on self-efficacy and presence. *International Journal of Human–Computer Interaction*, 37(2):181–196, 2021.

[22] Merel Krijn, Paul MG Emmelkamp, Roeline Biemond, Claudius de Wilde de Ligny, Martijn J Schuemie, and Charles APG van der Mast. Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour research and therapy*, 42(2):229–239, 2004.

[23] Fedwa Laamarti, Mohamad Eid, and Abdulmotaleb El Saddik. An overview of serious games. *International Journal of Computer Games Technology*, 2014, 2014.

[24] Pontus Larsson, Daniel Vastfjall, and Mendel Kleiner. Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments. Georgia Institute of Technology, 2001.

[25] Yun Ling, Harold T Nefs, Nexhmedin Morina, Ingrid Heynderickx, and Willem-Paul Brinkman. A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PloS one*, 9(5):e96144, 2014.

[26] Anita Dyb Linge, Stål Kapstø Bjørkly, Chris Jensen, and Bente Hasle. Bandura's self-efficacy model used to explore participants' experiences of health, lifestyle, and work after attending a vocational rehabilitation program with lifestyle intervention–a focus group study. *Journal of Multidisciplinary Healthcare*, 14:3533, 2021.

[27] Edwin A Locke. Self-efficacy: The exercise of control. *Personnel psychology*, 50(3):801, 1997.

[28] Benoît Massé, Silèye Ba, and Radu Horaud. Tracking gaze and visual focus of attention of people involved in social interaction. *IEEE transactions on pattern analysis and machine intelligence*, 40(11):2711–2724, 2017.

[29] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P Brooks Jr. Physiological measures of presence in stressful virtual environments. *Acm transactions on graphics (tog)*, 21(3):645–652, 2002.

[30] Isabelle Milleville-Pennel and Camilo Charron. Driving for real or on a fixed-base simulator: is it so different? an explorative study. *Presence*, 24(1):74–91, 2015.

[31] Chantal Olckers. Psychological ownership: Development of an instrument. *SA Journal of Industrial Psychology*, 39(2):1–13, 2013.

[32] Xueni Pan and Antonia F de C Hamilton. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3):395–417, 2018.

[33] Jorge Peña, Jeffrey T Hancock, and Nicholas A Merola. The priming effects of avatars in virtual settings. *Communication research*, 36(6):838–856, 2009.

[34] J Pettersson, A Albo, J Eriksson, P Larsson, KW Falkman, and P Falkman. Cognitive ability evaluation using virtual reality and eye tracking. In *2018 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*, pages 1–6. IEEE, 2018.

[35] Simon Riches, Soha Elghany, Philippa Garety, Mar Rus-Calafell, and Lucia Valmaggia. Factors affecting sense of presence in a virtual reality social environment: a qualitative study. *Cyberpsychology, Behavior, and Social Networking*, 22(4):288–292, 2019.

[36] Geneviève Robillard, Stéphane Bouchard, Stéphanie Dumoulin, Tanya Guitard, and E Klinger. Using virtual humans to alleviate social anxiety: preliminary report from a comparative outcome study. *Stud Health Technol Inform*, 154:57–60, 2010.

[37] McCartan Robson. *Real World Research*. Wiley, 2011.

[38] Daniel Roth, Jean-Luc Lugrin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*, pages 277–278. IEEE, 2016.

[39] Kerstin Ruhland, Christopher E Peters, Sean Andrist, Jeremy B Badler, Norman I Badler, Michael Gleicher, Bilge Mutlu, and Rachel McDonnell. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer graphics forum*, volume 34, pages 299–326. Wiley Online Library, 2015.

[40] Nicola S Schutte and Emma J Stilinović. Facilitating empathy through virtual reality. *Motivation and emotion*, 41(6):708–712, 2017.

[41] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[42] Tay Corrie Weber. Effects of own voice vs. another's voice during progressive relaxation. 1985.

[43] Brenda K Wiederhold. Can deepfakes improve therapy?, 2021.

Appendix A

Questionnaire

Dutch translation of the questionnaire. The original English version can be found in the work by Ding et al. [9].

Sense of Ownership
Q1: Ik voelde een sterke band met de stem die ik hoorde tijdens de sessie.
Q2: De stem die ik tijdens de sessie hoorde voelde aan als een deel van mijn eigen stem.
Q3: Ik voelde mij volledig op mijn gemak met de stem die ik hoorde tijdens de sessie.
Q4: Het voelde alsof de stem die ik hoorde tijdens de sessie "mijn eigen" gedachten waren.
Q5: Ik voelde een sterke link tussen mijzelf en de stem die ik hoorde tijdens de sessie.

Plausibility of the scenario
Q1: Tot op welke hoogte waren, de stem en de dialoog die je hoorde tijdens de sessie, natuurlijk en gepast tijdens je verkenning in de virtuele omgeving?
Q2: Tot op welke hoogte waren, de stem en de dialoog die je hoorde tijdens de sessie, een geloofwaardige afspiegeling van een vergelijkbare situatie in de werkelijkheid.
Q3: Tot op welke hoogte waren, de stem en de dialoog die je hoorde tijdens de sessie, specifiek en persoonlijk gerelateerd aan jou.

Appendix B

Script

The lines for Character A are lines of dialogue are recorded by the participant of the experiment. Character B is one of the virtual characters. Voiced by a female, and is not the participants. Character C is one of the virtual characters. Voiced by a male, and is not one of the participants.

B.1 Character A

Ik hoop dat ze begrijpen hoe lastig het is in deze markt. Ze moeten blij zijn als ik uberhaupt iets kan regelen.
Daar was ik al bang voor. Dit gaat erg lastig worden om iets te vinden.
Gelukkig hebben ze geen waanzinnige eisen. Ik denk dat ik wel wat voor ze kan vinden.
De weg kwijt zijn is nog zacht uitgedrukt. Ik kreeg bij het nakijken het idee dat ze helemaal niet hadden voorbereid.
Ik denk dat ze allebei wel de juiste intentie hadden. Wellicht dat ze het gewoon verkeerd hebben aangepakt.
Dit gesprek heeft me wel vertrouwen gegeven. Als ze slimmer te werk gaan, dan zullen ze beiden de cijfers nog wel kunnen ophalen.
Ik hoop dat ze weten waar ze aan beginnen. Deze hond is niet de makkelijkste. Ik zal moeten beoordelen of deze mensen geschikt zijn.
Dit is lastig. Ze lijken heel redelijk, maar als ze het heel druk hebben, dan is dit misschien geen goed idee.
Ze hebben mij toch overtuigt. Deze mensen lijken mij oprechte mensen. Als ze bereid zijn aanpassingen aan hun leven te maken, dan komt het vast goed.

Hallo en welkom. Wat fijn dat jullie er zijn.

Zoals jullie weten, zijn we hier vandaag om een promotie te bespreken. Jullie lijken beiden zeer geschikt, maar helaas kan er maar 1 de promotie ontvangen. Ik zal even uitleggen wat nu de bedoeling is.

Geef me een momentje.

Het is jammer dat ik ze niet allebei een promotie kan geven. Ik heb over beiden gehoord dat ze zeer capabel zijn.

Ik ga een aantal korte vragen stellen, die jullie dan beiden kunnen beantwoorden. Probeer het antwoord zo kort mogelijk te houden. Laten we beginnen met de eerste vraag. Waarom verdien jij de positie?

Ik heb veel positieve verhalen gehoord over jullie teams. Hoe schat je persoonlijk je vaardigheden in?

Jazeker u heeft mijn aandacht

U bent nog niet al te lang geleden gepromoveerd tot team leider. Hoe is het volgens u gegaan?

Ondanks haar leeftijd lijkt ze zelf verzekerd. Ik weet zeker dat ze ver zal komen.

U bent al langer in uw rol bezig. Wat heeft u geleerd over de jaren?

Bedankt voor het beantwoorden van mijn vragen. Jullie zullen beiden binnenkort weer van mij horen.

Wellicht kan ik de ander een opslag geven.

Zelfs als ze deze promotie niet krijgt, zal ik haar volgende jaar ongetwijfeld weer zien

Ik denk dat ik het nu wel weet. Het is verstandig om voor de meest ervaren kandidaat te gaan.

Waarom staat dit standbeeld hier op tafel? Kunnen ze dit niet beter ergens in een kast tentoonstellen?

Wat moet dat schilderij nou weer voorstellen. Ze hadden beter een schilderij over de natuur kunnen ophangen.

Ik hoop dat we snel klaar zijn.

Waarom staat dit standbeeld hier op tafel? Kunnen ze dit niet beter ergens in een kast tentoonstellen?

Wat moet dat schilderij nou weer voorstellen. Ze hadden beter een schilderij over de natuur kunnen ophangen.

Ik hoop dat we snel klaar zijn.

De kapstok is wel erg oudbollig zeg. Ze mogen hier wel eens de boel opknappen.

Ik heb niet veel tijd vandaag. Ik moet dit zo snel mogelijk afronden.

Dit kantoor is wel heel kaal. Ze mogen hier wel wat meer decoratie gebruiken.

B.2 Character B

Het is een leuke verassing om hier weer eens langs te mogen komen.
Mijn team heeft voor een langere tijd altijd in de top 3 gestaan. Ik denk dat dit aangeeft dat ik consistent kan presteren.
Sorry, heeft u het tegen mij?
Jazeker u heeft mijn aandacht
Ik heb over een lange carrière veel ervaring opgebouwd.
Ik heb niet de beste papieren, maar ik ben de meest ervaren kandidaat.
Ik heb geleerd hoe ik een team opzet dat langdurig productief kan zijn. Ook heb ik over de jaren voor heel veel verschillende problemen een oplossing gevonden.
Bedankt. Ik hoop snel van u te horen.
We zijn aangenaam verast dat we zo snel langs konden komen. Jullie zullen het in deze tijd wel erg druk hebben.
Het bevalt ons ook al een tijd niet meer, maar nu moeten we echt weg.
Ook zou een tuin heel erg fijn zijn. We zaten hiervoor in een appartement.
We zijn ook bereid om flexibel te zijn.
Ik denk dat ik het wel weet. Ik was totaal de weg kwijt tijdens het examen.
Hoe erg is het?
Ik had mij wel goed voorbereid, maar ik kon mij op de dag van het examen niets meer voor de geest halen.
Ik had me wel goed voorbereid. De vragen waren anders dan ik heb geoefend.
Dat kan het zijn. Ik heb inderdaag vooral gelezen en de lectures terug gekeken.
Bedankt voor het advies.
Wij kijken al een lange tijd uit naar vandaag. Een hond adopteren is iets wat wij al een tijd willen.
Ik ben dol op honden die er uit zien alsof ze nog in het wild kunnen rondlopen.
Ik heb nooit een hond gehad, maar heb er altijd al een gewild.
Nee, geen bijzondere reden. Ik heb het altijd druk gehad, dus het is er nooit van gekomen.
Wij zijn ook bereid om aanpassingen te maken aan ons leven. Ik ben ervan

overtuigd dat hij heel gelukkig bij ons zal zijn.
Zeker weten.
Wat een apart schilderij.
Wat is de betekenis van dit standbeeld?
Sorry, heeft u het tegen mij?
De kast bevat maar weinig interessante dingen.
Die palmboom doet me denken aan hoe hard ik een vakantie nodig heb.
Niemand weet wat dat standbeeld hier doet.
Die lamp wordt hopelijk binnenkort vervangen.

B.3 Character C

Bedankt voor de uitnodiging. Heel fijn dat er aan mij werd gedacht voor de promotie.
Ik verdien deze positie, omdat mijn team het best heeft gepresteerd vorig jaar.
Ik heb niet alleen de vereiste vaardigheden voor deze positie, maar ik heb ook de ambitie om nog verder te groeien.
Als team leider heb ik de afgelopen jaren een aantal succesvolle projecten voltooid.
Ik heb geleerd beslissingen te maken onder grote druk. Ook heb ik geleerd leiding te geven aan een groot team.
Bedankt voor deze kans. Ik waardeer het zeer.
Wij zijn op zoek naar een appartement met minstens 3 kamers en een grote keuken.
Verder is ons budget rond de 3 ton.
Wij hebben een beetje haast bij het vinden van een huis.
Ons huidige woonplaats moeten wij binnenkort al verlaten.
De locatie is niet ideaal. Wij zouden graag dicht bij een school wonen.
Ik hoop snel weer een huis te bezichtigen.
Ik ben erg benieuwd naar hoe ik het heb gedaan.
Ik heb me niet zo goed voorbereid als gepland.
Ik snap niet hoe dit fout is gegaan.
Wat kan ik volgende keer beter doen?
Bedankt voor alle feedback. Ik weet nu wat ik moet doen.
De hond zag er mooi uit op de foto's. Wij zijn al een tijdje op zoek.

Ik heb in mijn jeugd ook een hond gehad.
Ik heb interesse in Duitse Herders.
Ik werk veel thuis, dus het uitlaten is geen enkel probleem.
Als het nodig blijkt te zijn, kunnen wij een trainer inhuren.
Ik eet echt niet veel meer dan de anderen.
Ik eet veel vlees, dus dat is toch niet slecht.
Ja maar ik sport heel veel om het te compenseren.
Ik blijf maar aankomen. Ik snap er niks van.
Moet ik dan op water en brood gaan leven?