Enabling joint exploration of virtual environments for treatment of substance use disorders using an asymmetric virtual reality approach

W.S. Ammerlaan

June 16, 2022

M.Sc. Interaction Technology

Human Media Interaction Group

Faculty of Electrical Engineering, Mathematics & Computer Science

University of Twente

Author

Wessel S. Ammerlaan

Examination committee

dr. ir. Randy Klaassen
Human Media Interaction Group
University of Twente

prof. dr. Dirk Heylen
Human Media Interaction Group
University of Twente

dr. ir. Robby van Delden
Human Media Interaction Group
University of Twente

Jan Kolkmeier, MSc
Human Media Interaction Group
University of Twente

Simon Langener, MSc
Human Media Interaction Group
University of Twente

dr. ir. Wouter Eggink
Interaction Design Group
University of Twente

June 2022

Acknowledgements

First and foremost I thank my daily supervisors, Randy, Jan, and Simon who put in the time and effort to support me throughout the full length of this thesis. Although our meetings could be confusing or slightly frustrating at times, I learned much from your insights and you were able to motivate me to go beyond what I was capable of at the time. There was always room for a laugh as well and I think that anyone can be happy to have you as supervisors.

Furthermore, the help from Joanneke, Laura, and Leonie from Tactus cannot go unmentioned. Specifically, I thank Laura and Leonie for the effort they put in such that I could run my evaluations on the clinic with actual treatment providers and I thank Joanneke for the insights that complemented those of Randy and Simon on topics related to psychology, a field that I was relatively unfamiliar with before the start of this thesis.

Then to express my appreciation for Thérèse, who was at my side, often literally, the past years including all its ups and downs.

Lastly, I want to thank my parents for always being supportive throughout my years at university and the help making this chapter of my life a reality in the first place.

Abstract

When using immersive virtual reality (IVR) technology, such as head-mounted displays (HMDs) to visit virtual environments (VEs), exploration of the VE in both a navigational and utilization sense is key to gain an understanding of that VE. Exploring together with others, i.e. joint exploration, provides collaborative opportunities which can benefit the exploration process. In this work we focus on joint exploration of VEs using IVR for substance use disorder (SUD) treatment, including treatment of people with mild intellectual disability or borderline intellectual functioning (MBID; IQ 50-85). At clinics, IVR can be used to let clients with SUD experience ecologically valid scenarios in order to determine what elements in those scenarios relate to their SUD, why those scenarios might be problematic, and how to cope with problematic scenarios. Exploring these situations together with treatment providers allows clients to discuss their experiences and emotions in the moment. However, the closed-off nature of the HMD hinders communication and diminishes the benefits of using IVR. In this research, we aimed to determine how to facilitate communication for joint exploration to support SUD treatment of clients with MBID using an asymmetric virtual reality (AVR) approach. AVR combines IVR and less immersive displays to provide multiple users access to the same shared VE. Letting the treatment provider use a less immersive display for joint exploration helps them ensure the safety of the client who is wearing the HMD in real life during the shared experience. For this aim, we executed two user-centered design iterations. The first iteration resulted in a general desktop-based approach for the treatment provider to use. Having a research assistant role-play a client, this approach was evaluated with five treatment providers to determine how to specialize the approach further for SUD treatment. In the second iteration the desktop was substituted for an approach involving a tablet and IVR position trackers. The second iteration approach was evaluated with three pairs of treatment providers with one of each pair role-playing a client. Both evaluations consisted of exploration exercises with elements of SUD treatment. Over the course of these two iterations we found that for SUD treatment it is integral that an approach should not hinder the client in obtaining and maintaining presence, users can create a mutual understanding using verbal communication and creating awareness of the visual perspective of the client, virtual representations of treatment providers cannot display inappropriate behaviour, and that a solution should not rely on physical interaction. Furthermore, we found that the final tablet-and-tracker-approach was able to facilitate communication for joint exploration on a practical level by enabling communication using a combination of verbal, gestural, and graphical cues. However, with regard to embodied interaction, the approach lacked in its human qualities, ability to support communication on affect, and the use of less immersive technology sat in the way of the naturalness of the interaction. Moreover, the shared physical space plays a substantial role considering shared movement in the VE, treatment providers wanting to distance themselves from the VE, potential group sizes larger than two, and practical limitations of a system's spatial footprint. This work seems to be the first to combine a focus on joint exploration using AVR and SUD treatment.

Contents

Abbreviations and acronyms

AR - augmented reality
AST - asymmetric virtual reality for substance use disorder treatment
AUD - alcohol use disorder
AV - augmented virtuality
AVR - asymmetric virtual reality
MBID - mild intellectual disability or borderline intellectual functioning
CAVE - cave automatic virtual environment
CBT - cognitive behavioural therapy
CGT+ - Cognitieve Gedragstherapeutische behandeling Plus
CVE - collaboration in virtual environments
EPU - externally partaking user
EQ - evaluation question
FIE - first iteration evaluation
FOV - field of view
FR - functional requirement
HMD - head-mounted display
ID - identification
IQ - intelligence quotient
IVR - immersive virtual reality
IVRU - immersive virtual reality user
JN - joint navigation
MDOD - Minder Drank of Drugs
MR - mixed reality
NFR - non-functional requirement
PC - personal computer
PI - place illusion
PMR - pure mixed reality
POF - point of focus
POV - point of view
Psi - plausibility illusion
RQ - research question
SQ - sub-question
SUD - substance use disorder
TM - tracked mode
UCD - user-centered design
UI - user interface
UM - untracked mode
UR - user requirement
VE - virtual environment
VR - virtual reality
VRET - virtual reality exposure therapy
WYSIWIS - what you see is what I see
XR - extended reality

1 Introduction

Immersive virtual reality (immersive VR or IVR) technology, such as head-mounted displays (HMDs), provides various opportunities for visiting known or new locations. Visiting these locations can be as simple as putting on an HMD and stepping into the virtual world. Hence, such virtual visits are less bound to physical, financial or time constraints compared to real life. Accessible destinations can range from replicas of existing locations to historically accurate replicas to completely fictitious locations. IVR technology is applicable in a variety of sectors, such as entertainment (*e.g.* VR escape rooms and VR arcades), travel (*e.g.* VR tours), real estate (*e.g.* property viewing), and healthcare (*e.g.* VR therapy and VR getaways for people who cannot leave their beds).

One aspect that is relevant for any virtual visit is that upon entering a virtual environment (VE), there is a need to explore that environment in order to learn what a VE has to offer in terms of what there is to see, but also in what there is to do. As such, the notion of *exploration* encompasses both an aspect of *navigation*, for learning about the spatial aspect of a VE, and *utilization*, for learning about the functional aspect of a VE. Because VEs can be vastly different from one another or real life in one or both of these aspects, knowing the implications of aspects of one VE does not guarantee that these implications translate over to other VEs as well. Therefore, the exploration process is often a necessity when entering an unknown VE. For example, the VE might contain a virtual zoo which only contains imaginary creatures or the VE is a replica of a convention stage to practice public speaking.

While it is often possible to explore VEs alone, there might be a need or desire to explore together with others, which is referred to as *joint exploration*. The reason to explore together with others, and who those others are, often depends on the users themselves and the application of the VE. For example, someone might not want to explore alone because they find it more enjoyable or comforting to go together or someone might only be allowed to explore together because they need parental or professional supervision. Regardless of the reason, joint exploration provides collaborative opportunities which can benefit the exploration process for both the navigation and utilization aspect, as users can share with the other what they have already learned. For example, helping each other by highlighting important spatial elements that the other might have missed or discussing how to best take on a social encounter presented in the VE. However, there are two challenges that need to be addressed before joint exploration using IVR can take place.

1.1 Challenges and scope

The two challenges that need to be addressed in order for joint exploration using IVR is able to take place are: enabling a multi-user VR experience and enabling communication between those users. In this research we will focus on asymmetric virtual reality (AVR) to enable a multi-user VR experience. In AVR, IVR technology is mixed with less immersive displays to provide multiple users access to the same shared VE [66]. The asymmetric approach allows multiple people to explore a virtual reality together without the need for as many IVR systems, compared to approaches that require each user to have their own individual IVR setup, whilst being more flexible in how the experience is provided to users. The latter is important as it allows for people that cannot or do not want to make use of IVR technology for any reason to partake in shared VR experiences as well. An example of an AVR system is shown in Figure 1.

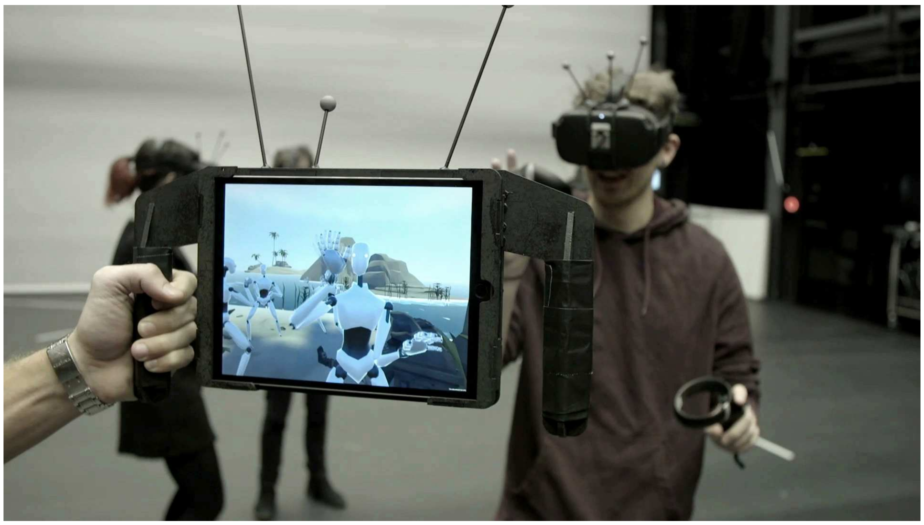


Figure 1: Example of an asymmetric virtual reality system. Several people are immersed in IVR by means of HMDs and one observes the virtual reality via a portable display. In the virtual environment, the IVR users are represented via humanoid avatars that supposedly move according to the position of the HMD and controllers. It is clear that the non-VR user is represented in some way, as the IVR user is seemingly able to acknowledge their presence. However, it is not clear whether it is the same way the IVR users are represented. Photo by Stefan Dux © ZHdK 2020 from `https://blog.zhdk.ch/immersivearts/multi-user-vr/` [visited 14-March-2022].

However, due to the combination of IVR and non-IVR technology, AVR cannot provide the exact same experience to all users simultaneously. What a suitable or desired joint exploration experience is depends considerably on the application for which the AVR system is used [68], including factors such as the number of users and co-location or dislocation of those users, which in turn substantially influences what is required in terms of communication between users as well. For example, the IVR users in Figure 1 are represented by virtual embodiments, *i.e.* avatars, which hands and body move according to the position of the HMD and controllers of the user, making it suitable to look and point to objects in the VE. However, it is not clear if the non-immersed user can only observe or is able to point to objects in the VE as well in some way.

Covering all potential applications and how they might affect communication is not doable within a single research, hence in this research we address joint exploration using AVR within a single use case. Similarly, we can evaluate an envisioned solution with only a limited number of people within a given time-frame, which provides us with results of a limited representation of the target group. However, this does allow us to obtain ground to reason about the implications of an envisioned solution for the specific use case, which can then be generalized to implications for the general case of joint exploration using AVR. This process of tightening the scope and generalizing results is common in research and can, to certain extent, be visualized as an hourglass. A visualization of the hourglass in context of this research is shown in Figure 2. The use case that will be covered by this research regards mental healthcare: using AVR to support treatment of people with a substance use disorder (SUD) at Tactus.

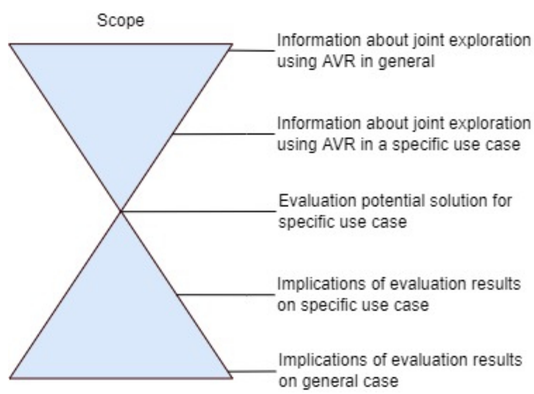


Figure 2: Hourglass visualization of tightening the scope and generalizing results thereafter for this research. The scope is tightened from the general case of joint exploration using AVR to a specific use case application up to an evaluation with a limited number of people from the target group. After the evaluation, the implications of its results can be discussed specific to the use case and thereafter to the general case.

1.2 Use case

This research will focus on using AVR to support treatment of people with a substance use disorder (SUD) at Tactus. Tactus is a Dutch mental healthcare organization, specialized in treating various forms of addictions amongst which is SUD. SUD is the persistent use of substances, such as alcohol or cannabis, despite harmful consequences.

A portion of Tactus' clients are people with mild intellectual disability or borderline intellectual functioning (MBID; IQ 50-85), characterized by significant limitations in both intellectual functioning and in adaptive behaviour, which cover many everyday social and practical skills [77]. People with MBID have been identified as a risk group for more severe negative consequences of substance use and for developing SUD [92] and considering that this group is impulsive, impressionable, has a hard time generalizing, and seeing cause-effect relations, treatment adapted to better match these characteristics is therefore required [93].

In general, a common barrier to successful treatment outcomes for SUD is relapse [51, 103] and this is where IVR comes into play. Relapse is tempted by various *high-risk* situations, such as the desire to use a substance, *i.e.* craving, or specific social contexts that make it harder to refrain from the use of substances [102]. IVR can be used to elicit or simulate these high-risk situations tailored specifically towards what is relevant for a client. Using IVR, these high-risk situations can be explored to determine which situations might be problematic, what makes a situation problematic, determine which coping strategies could be used, and practice these coping strategies in the moment [78]. This makes it more likely for results obtained during practice to be carried over to real life [9], *i.e.* having a higher ecological validity.

Next to being more ecologically valid, using IVR to provide such high-risk situations comes with various more benefits. A summary of the benefits provided by Botella *et al.* [9] and Riva [74] is: the availability of situations, relatively low cost of providing the situations, having total control over the situations, being able to provide the situations in the privacy and confidentiality of a clinic or office, providing context for clients who have trouble imagining, provide situations that go beyond what can be experienced in real life, and IVR allows for the construction of scenarios in which the users experience themselves as competent and efficacious with to make them realise that difficult situations can be overcome through confrontation and effort.

Learning what situations are relevant to someone regarding SUD and how they are relevant, as well as learning and trying out various coping strategies and seeing how they affect a situation or oneself in that situation, are all examples of the utilization aspect of exploration relevant to the use case. However, that does not make the navigational aspect of exploration less important. For example, if a client has a hard time pinpointing what situations are risky for them, navigating through a larger VE containing various SUD-related encounters might help them identify problematic elements. Or, if a coping strategy relates to a specific object, a client might need to locate that object first before being able to apply that strategy.

Making exploration of these situations a joint exploration experience, with client and treatment provider together, allows for them to directly exchange emotions and experiences while they are happening. It allows the treatment provider to support the client through these emotions and experiences in the moment as well. For example, by helping with determining possible coping strategies or role-playing a character relevant to the client.

Using an AVR approach for SUD treatment seems a good fit as well, as this would allow the treatment provider to keep an eye on the client in real life to ensure the client's safety during the experience. For example, ensuring the integrity of the room or intervening when a client is about to bump into something. Regarding the challenges of enabling joint exploration, the question that remains is *how* to facilitate communication for joint exploration to support SUD treatment using AVR.

1.3 Research questions

The challenge of enabling joint exploration is twofold by the need for providing a multi-user IVR experience and enabling communication between those users in a way that is substantially influenced by the application for which it is used. In this research we will focus on taking on these challenges in the context of IVR for SUD treatment and use AVR to provide the multi-user experience. The question of how this AVR solution should facilitate communication in the context of IVR therapy for SUD is what we set out to answer in this research and is captured in the main research question:

> **RQ1: How to facilitate communication for joint exploration to support SUD treatment of people with MBID using asymmetric VR?**

To make a better attempt at answering RQ1, we first need to take a closer look at the critical implications the application of IVR for SUD treatment has on an envisioned solution, such that it can be taken into account from the start. Not taking these implications into account may result in a solution that enables joint exploration but is ultimately unsuitable to be used in the context of SUD treatment. Determining what these critical implication are is captured in SQ1:

> **SQ1: What are critical aspects that an AVR solution needs to adhere to to make that solution suitable for use in SUD treatment?**

Additionally, we need to take a closer look at what is necessary for effective communication on both the joint navigation and collaboration aspect of joint exploration to enable that communication altogether. Determining what is necessary for communication on both aspects is captured in SQ2 and SQ3:

> **SQ2: What is necessary for effective communication for joint navigation?**

> **SQ3: What is necessary for effective communication for collaboration in virtual environments?**

1.4 Approach and outline

To attempt to answer the main research question, we took a user-centered design (UCD) approach with two iterations. In the first iteration we started out by building an initial design rationale by provisionally answering SQ1 through SQ3 using a mostly theoretical approach as well as looking at related research and systems with modest involvement of SUD treatment experts (Chapter 2).

Using this design rationale we set up system requirements followed by ideation resulting in a first prototype design (Chapter 3). After implementation and refinement, this first prototype is then evaluated on how it can be adapted to be a better fit for use for SUD treatment (Chapter 4), ending the first UCD iteration. In the second iteration, we start out with ideation on how the prototype could be adapted to incorporate the insights provided by the first iteration evaluation, resulting in a second version of the prototype. This second iteration prototype is in turn evaluated to gain the necessary insights to attempt to further answer SQ1 through SQ3 and ultimately answer RQ1 (Chapter 5).

The results of both evaluations allowed us to determine that using the mostly theoretical design rationale and specializing a resulting prototype in a follow-up iteration is a viable approach for creating a solution that is effective on a practical level for enabling communication on joint exploration. However, evaluation within the SUD treatment context showed a necessity to go beyond the practical and make the communication between users more personal with regard to embodied interaction, highlighted the substantial role the shared physical space plays, as well as additional system constraints SUD treatment brings forward that were not yet taken into account. These aspects are to be considered for joint exploration using AVR in general as well, next to the desired degree of presence experienced by the non-immersed user and what other actions an application requires next to those required for joint exploration. These findings as well as the limitations of our work and directions for future work are discussed and concluded upon (Chapter 6). In terms of the hourglass analogy, this approach is as shown in Figure 3.

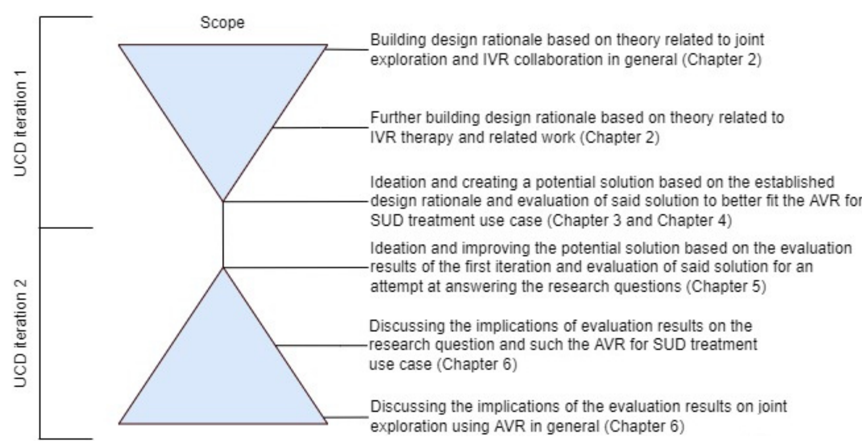


Figure 3: Hourglass visualization of tightening the scope and generalizing results thereafter and how that analogy is reflected in the structure of the approach of this research specifically.

role and the other had to act like a SUD client in two exercises that focused more on navigation and the therapeutic aspect respectively. After the exercises, the experts were asked about their experiences in a focus group.

2 Background

In this chapter we create an initial design rationale by provisionally answering SQ1 through SQ3 as well as looking at related work. In Chapter 2.1 information on SUD treatment and the implications of IVR and AVR on treatment is provided for SQ1, followed by theory on navigation and joint navigation for SQ2 in Chapter 2.2 and theory on collaboration in VEs for SQ3 in Chapter 2.3. Then we take a closer look at related work in Chapter 2.4 and discuss our findings in relation to SQ1-3 in Chapter 2.5. Implications of these findings on ideation and design are discussed in the next chapter, Chapter 3.

Information on navigation and collaboration in VEs was obtained in previous work on group navigation in IVR [2]. A selection of that information relevant for this research is presented in Chapter 2.2 and Chapter 2.3 respectively.

2.1 Asymmetric virtual reality for substance use disorder treatment

Before we can determine what the critical aspects are that an AVR solutions needs to adhere to to make a solution suitable for SUD treatment, we need to establish what a SUD is, what current treatment approaches are and how Tactus incorporates those approaches to determine why IVR for treatment seems promising. Then we can determine what IVR entails, how treatment can benefit from it and what the results of current applications of IVR for SUD treatment are. Thereafter we can determine the critical aspects of IVR to obtain those results as well as the potential of AVR and the role of joint exploration within IVR treatment.

2.1.1 Substance use disorder and treatment at Tactus

Substance use disorder (SUD) is the persistent use of substances, *i.e. drugs*, despite harmful consequences, with addiction being the most severe variant of the disorder.[1] Commonly known examples of such substances are alcohol, nicotine, marijuana and cocaine. Getting or maintaining an SUD is influenced by a variety of factors, hence there are various approaches on how to deal with SUDs. As described by Kerssemakers *et al.* [48], these approaches can be either preventive or curative. In preventive approaches, the use of a substance is prevented or discouraged, for example by making the substance illegal or prosecuting substance users. In curative approaches, a substance user receives treatment to stop or deal with substance use, for example by learning how to stay abstinent or by receiving medicinal treatment focused on the brain. In this research, the focus lies on curative approaches, specifically the behavioural therapeutic approach. In this approach SUD is seen as learned behaviour, where the positive effects associated with the substance use keep a person addicted. The treatment ap-

[1] Help With Addiction and Substance Use Disorders by American Psychiatric Association: `https://www.psychiatry.org/patients-families/addiction` [Accessed January 18, 2021]

proach presumes this learned behaviour can also be reversed by learning other associations.

A portion of Tactus' clients are people with mild intellectual disability or borderline intellectual functioning (MBID; IQ 50-85), characterized by significant limitations in both intellectual functioning and in adaptive behaviour, which cover many everyday social and practical skills [77]. People with MBID have been identified as a risk group for more severe negative consequences of substance use and for developing SUD [92] and considering that this group is impulsive, impressionable, has a hard time generalizing, and seeing cause-effect relations, treatment adapted to better match these characteristics is therefore required [93].

Tactus offers two treatment protocols specifically focused on people with MBID: *Minder Drank of Drugs* (MDOD) [93] and *Cognitieve Gedragstherapeutische behandeling Plus* (CGT+) [91]. Both protocols contain many of the same elements in their content with the main differences between MDOD and CGT+ being that MDOD also includes group sessions and the total number of sessions is higher. MDOD specifically uses limited vocabulary, repetition and use of game elements, keeping in mind the traits of the target group. The CGT+ protocol is a variation of the regular cognitive behavioural therapeutic protocol (abbreviated CGT in Dutch and CBT in English), adapted to help people with MBID. The original CBT protocol has only one longer meeting once a week, whereas CGT+ splits this into two separate shorter meetings to ensure that the client can stay focused and allows for repetition.

A core part of both programs is identifying the risk level of various situations and discussing self-control techniques for how the client could deal with these situations. Situations are grouped into three categories with respect to how risky they are for resulting in substance use and are given the following corresponding colours: green for no/low risk, orange for when needing to be cautious, and red for high risk. Examples of risky situations are the desire to use a substance, *i.e.* craving, or specific social contexts, such as attending a football game, that make it harder to refrain from the use of substances [102]. The self-control techniques are presented as a mnemonic called the 6Ds. Table 1 provides an overview of the 6Ds as well as the situation category in which they are relevant, technique type, and examples.

A common barrier to successful treatment outcomes for SUD is relapse [51, 103] and this is where IVR comes into play. Relapse is tempted by the aforementioned risky situations, such as craving or a specific social context [102]. IVR can be used to elicit or simulate these risky situations, tailored specifically to be meaningful and relevant to a client, through realistic environments and allowing for interactive behaviour [87]. Using IVR, these risky situations can be explored to determine how a situation makes the client feel, what makes a situation problematic, and how to deal with the situation [78], which makes it more likely for results obtained during treatment to be carried over to real life [9]. The concept of treatment results carrying over to real life is known as *ecological validity.* Thus IVR is a promising tool for supporting current treatment approaches or even go beyond them in more novel approaches, especially con-

Table 1: Self-control techniques: 6Ds [93]

Technique	**Situation**	**Example(s)**
Distance	Risky (red & orange)	Going for a short walk
Distraction	Risky (red & orange)	Talking about or doing something else
Declare	Risky (red & orange)	Expressing what you (do not) want and calling for help
Different thinking and different acting	Risky (red & orange)	Thinking about the consequences of use, choosing a non-alcoholic drink at a football game or birthday
Doing great	Before and after (green)	Rewarding desired behaviour
Deals	Before and after (green)	Making rules (deals) about limits of use and consequences of undesired behaviour

sidering the implications of MBID on treatment approaches. However, *virtual reality* is an ambiguous term in research and therefore it is necessary to establish what VR can be and why we use the term IVR for this research.

2.1.2 Virtual reality and immersive virtual reality

Virtual reality (VR) is widely used as an umbrella term for all simulated experiences that try to substitute the real world with a completely virtual one. Milgram and Kishino [58] suggest that reality and virtual reality are both ends of a continuum, with mixed reality (MR) in between. Mixed reality consists of augmented reality (AR) and augmented virtuality (AV) that meet on an undefined point. An illustration of the continuum with visual examples is found in Figure 4. Flavián *et al.* [28] suggested to adjust the continuum by including pure mixed reality (PMR) in the middle of the continuum to better delineate the terms AR, AV, and MR, which is determined to be used interchangeably in research. With this adjustment, AR refers to experiences where virtual elements solely overlap the real world, AV refers to experiences where real world elements solely overlap the VE, and PMR refers to experiences where virtual elements are merged into the real world such that both virtual and real elements can interact in real-time. A term that gained more traction in recent years and encompasses all of MR and VR is extended reality (XR) [34].

A conventionally held view of VR is that the user is totally occluded from the real world in order to experience a virtual one, *i.e.* the user is totally immersed by the VR technology. However, the term is also used for systems that do not match this view [58]. For example, using a standard television and a balance board [57] or referring solely to a virtual world, such as Second Life [85]. Due to this ambiguity, we resort in this research to the term *immersive virtual reality* (IVR), which is in line with the conventionally held view of VR, but emphasizes

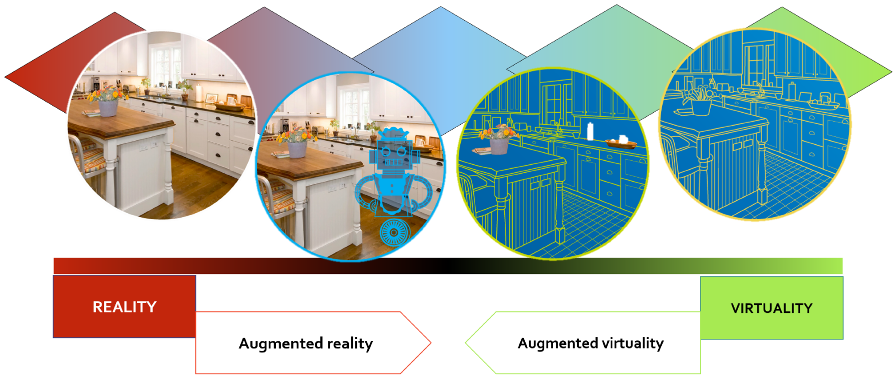


Figure 4: Illustration of the virtuality continuum as described by Milgram and Kishino [58]. Virtuality is a dated term for what is now more commonly referred to as virtual reality. Illustration by Sara Muršić from `https://augmentedtoys.org/augmented-reality/`.

the use of highly immersive technology, such as head-mounted displays (HMDs) [31] or cave automatic virtual environments (CAVE) [20]. HMDs occlude the users visual contact with the real world and replace it with a computer-generated image stream, which is dynamically adapted to the viewing position of the user and often includes headphones to include sound from the virtual world and to block sound from the real world. A CAVE is a room-sized area which surrounds a user with multiple stereoscopic projections and surround-sound speakers that adapt to the viewing position of the user as well. Visual examples of an HMD and CAVE are shown in Figure 5.

The simulated scenario that is presented to (I)VR users is commonly referred to as the virtual environment (VE), in which the user can navigate and interact with the objects that it contains. Input devices allow users to interact with the virtual environment by tracking the user's actions and sending them to the computer. Examples of such actions are head, limb, and hand movements that are recorded using head-positioning sensors and hand-held controllers. Whenever two or more users inhabit the same VE, there is often the possibility for communicating and collaboration by means of avatars, which are virtual embodiments of the users within the VE that are directly controlled by those users in real time. VEs may include other social entities that are not controlled by other users, but by the computer itself. If computer-controlled, these entities are referred to as embodied virtual agents, or agents for short, rather than avatars [67].

2.1.3 IVR for SUD treatment

As mentioned earlier, IVR technology is a promising tool for supporting current or novel treatment approaches due to its ability to elicit or simulate situations

Figure 5: Visual examples of an HMD (left) and a CAVE (right). Photo HMD by Burst (@Burst) from `https://www.pexels.com/photo/man-wearing-black-vr-goggles-373905/` [visited 01-April-2022] and photo CAVE by Dave Pape from `https://commons.wikimedia.org/w/index.php?curid=868395` [visited 01-April-2022]

tailored specifically to be meaningful and relevant for a client through realistic environments and allowing for interactive behaviour [87], making treatment more likely to be ecologically valid [9]. More specifically, as Teo *et al.* [87] explains, IVR providing these meaningful and interactive environments capitalizes on neuroplasticity, which is the ability of neural networks in the brain to learn or extinct conditioned responses.[2] For SUD these conditioned responses refer to the brain's reward system reflexively responding to any cue related to a substance, leading to a drive toward substance use [64]. This conditioned response is referred to as cue reactivity and relates to craving, but also psychophysiological responses, such as skin conductance, heart rate, and temperature [17].

Next to being able to capitalize on neuroplasticity, IVR's ability to provide meaningful and relevant experiences for therapy translates into various other benefits. A combined summary of relevant benefits as described by Botella *et al.* [9] and Riva [74] is as follows:

- *Communication medium*: IVR is able to generate experiences that provide new knowledge and meaning related to users' problems. For example, by showing how users perceive the world and how their view does not match with the real world.
- *Realism and presence*: when users judge their situation as real, they behave like they would have if the situation would occur in real life. This is particularly useful for clients who have trouble imagining.
- *Ecological validity*: for CBT and similar strategies, IVR can be used to implement the relevant techniques more ecologically valid, meaning that

[2]Neuroplasticity in Wikipedia: `https://en.wikipedia.org/wiki/Neuroplasticity` [Accessed 21-March-2022]

these techniques can better elicit cue reactivity and results are more likely to be carried over to real life. For example, by directly applying coping strategies when exposed to critical stimuli, such as social situations, in IVR instead of when they occur on a later point in real life.

- *Going beyond reality*: IVR makes it possible to provide situations or elements that go beyond what can be experienced in the real world.
- *Perceived self-efficacy*[3]: IVR allows for the construction of scenarios in which the users experience themselves as competent and efficacious with the goal that difficult situations can be overcome through confrontation and effort.
- *Control and safety*: there is total control over every element in the simulated experience, which can be adjusted according to a user's need, additionally providing a sense of safety for clients as there are no unknown variables compared to the real world.
- *Privacy and confidentiality*: therapy techniques can be applied within the therapist's office without judgement of the outside world.
- *Availability*: critical scenes and scenarios are always accessible during therapy and can be used at the time or rhythm a client wants or needs.
- *Cost*: creating virtual experiences is relatively cost efficient compared to creating similar experiences in the real world.

Of course, these benefits are only applicable if treatment results using IVR are on par or better than conventional approaches. For treatment of anxiety disorders, such as specific phobias, social anxiety, and post-traumatic stress disorder, IVR exposure therapy (VRET) seem to have indicated its worth [10, 74]. In exposure therapy, clients are exposed to the things they fear or avoid in order to remove the conditioned psychological response[4] and in VRET, IVR technology is used to provide this exposure. Results of using VRET are on par with using traditional methods compared to control conditions [10, 74] and these results are relatively consistent across the aforementioned disorder types [16]. However, treatment for SUD based exclusively on VRET to substance-related cues produced mixed results [78, 90]. As indicated in a recent review on the use of IVR for SUD treatment by Langener *et al.* [51], the clinical value of IVR in assessment and treatment of SUD is yet to be backed by a substantial body of research, but current results show indeed that IVR for SUD seems to find its strength in providing ecologically valid environments and eliciting cue reactivity. Thus IVR for SUD treatment can be helpful in approaches that focus

[3]Perceived self-efficacy as proposed by Bandura [5]: judgement of how well one can execute courses of action required to deal with prospective situations.

[4]What Is Exposure Therapy by American Psychological Association: `https://www.apa.org/ptsd-guideline/patients-and-families/exposure-therapy` [Accessed January 18, 2021]

on neuroplasticity such as learning and applying coping strategies directly in the VE. Considering the importance of providing ecologically valid environments and eliciting realistic responses such as cue reactivity in clients, we shift our focus to the experience of *presence* when using IVR technology, to understand better *why* IVR is able to provoke such responses.

2.1.4 The role of presence in IVR for SUD treatment

The feeling of being inside the virtual world is regarded to as sense of presence, also referred to as *spatial presence* or simply presence. Similar to the team behind the igroup presence questionnaire [44], we make the distinction that spatial presence is the subjective sense of being in a VE, whereas *immersion* is the objective ability of the technology to replace the real world with a virtual one. The rationale for the desire of high presence for IVR applications in research, is that the user should behave and react in the same way in the virtual scenario as in reality.

In literature, the definition of *presence* has been debated, evolved, and ramified depending on the context in which it is used. In Coelho *et al.*'s [19] attempt at defining presence, a distinction is made between *media presence* and *inner presence*, where it is argued that they are two sides of the same coin. Media presence is described as *the perceptual illusion of non-mediation*, *i.e.* presence is achieved whenever people fail to perceive or acknowledge the existence of a medium such as IVR in how they perceive the VE and respond as if the medium would not be there. Media presence is in that sense relatively predictive: to increase presence, experienced mediation must be reduced, in which the technology's immersive capabilities play a crucial role [74]. For the second distinction of inner presence is argued that presence is the outcome of a simulative phenomenon in the brain that is not necessarily linked to a medium, as described by Riva [74]. As Riva explains, this simulative phenomenon refers to an internal model by the brain of both the body and its proximity, which is used to provide predictions about expected sensory input and to minimize the number of prediction errors (*i.e.* surprises). They argue that inner presence is related to this brain model in the sense that the more similar a VE conforms with the brain model, the more the user feels present in that VE. Put simply for IVR, IVR technology attempts to predict the sensory consequences of users' actions by showing them the same outcome expected by our brains.

Slater [82] provides additional insight on why people respond realistically to IVR. They argue that next to feeling presence in a VE, there is also the need for people to believe that what is happening in that VE is actually happening in relation to them. They support this by introducing the concepts of place illusion (PI) and plausibility illusion (Psi). PI is similar to the notion of spatial presence as it refers to the sense of *being there*, but excludes other multiple meanings that have been attributed to the word presence. Psi on the other hand is determined by the extent to which the system can produce events that directly relate to someone and therefore also the extent of the overall credibility of the scenario being depicted in comparison with a person's expectations. A key component of Psi is that events in the virtual environment over which you have no direct control refer directly to you, which does not require physical realism. For example, another user or agent reacting to a user's proximity would be such an event and would therefore help obtain and maintain Psi. They argue that when both PI and Psi occur, people will respond realistically to virtual reality, despite those people knowing that they are not actually in the VE and that events there are not actually occurring.

Considering the importance of presenting ecologically valid environments and eliciting realistic responses for SUD treatment, a relation between presence, in any form, and SUD treatment outcomes appear to be relatively understudied. Two connected studies that considered relation of spatial presence and cue reactivity, positively correlated presence and craving levels of smokers [26, 27]. A more recent work by Simon *et al.* [80] regarding alcohol found that craving levels more strongly increased with perceived ecological validity in heavy drinkers than in occasional drinkers. They considered perceived ecological validity a part of presence together with spatial presence, the feeling of being psychologically involved similar to Psi, and any adverse physiological reactions. For these latter three dimensions, no noticeable effect was found in their evaluation. Note that the aforementioned adverse physiological reactions, amongst others, refer to VR sickness.

VR sickness, also called cybersickness, occurs whenever VR technology tries to visually mediate self-motion but fails to mediate other organs related to motion, such as the vestibular system [52], resulting in symptoms such as headaches, nausea, fatigue, and disorientation [101], and is negatively correlated with the feeling of presence [97]. In contrast to SUD treatment, spatial presence is widely regarded as a necessary mediator for VRET that allows real emotions, such as fear, to be activated by a virtual environment [23] and is seemingly necessary for successful outcomes of therapy [62]. However, a causal relationship has not been demonstrated yet [36].

The information provided up until now was mainly concerned with IVR users and clients using IVR in the context of SUD treatment. What is left to determine is the role of the therapist in this IVR experience, how AVR supports this role, and the implications on presence.

2.1.5 AVR for SUD treatment

Similar to the importance of establishing a concrete understanding of the distinction between VR and IVR made in this research, we first need to establish what AVR entails. As per Ouverson *et al.* [66], AVR is a form of MR which merges non-immersive displays with immersive HMDs to facilitate access to a shared VE. They note that, despite that asymmetric use of VR has been studied for longer under the term MR or mixed-space collaboration, the term *asymmetric VR* is relatively new. They explain that the term essentially refers solely to the use of VR by a *co-located* group of individuals, however, they determine that the term appears in works regarding similar applications for spatially separated users as well, such as the work of Jeong *et al.* [45]. This is also reflected in the definition provided by Horst *et al.* [42] who suggests that AVR is a subclass of multi-user VR that offers not all participants the same interaction possibilities with the VE, leaving out location of users as well. Ouverson *et al.* [66] adds that much academic effort went into similar applications for spatially separated users, leaving AVR with co-located users relatively unexplored. In this research, our main focus is on AVR with co-located users and as such when referring to AVR it refers to AVR with co-located users, unless explicitly stated otherwise.

As explained earlier in Chapter 2.1.3, IVR for SUD seems to find its strength in providing ecologically valid environments and eliciting cue reactivity, which are helpful in approaches that focus on neuroplasticity, such as learning and applying coping strategies directly in a relevant context provided by the VE. AVR comes into play when it is deemed beneficial or necessary for the therapist to help the immersed clients without interrupting the IVR experience. IVR technology's goal is to occlude as much of real life as possible, which includes the non-immersed therapist, making it substantially harder for the therapist to support a client throughout their IVR experience. Using AVR, the therapist can also be provided access to the VE and as such be included in the IVR experience with the client.

When providing an AVR experience, the notion of *social presence* becomes relevant as well. Social presence is the experience of being with another, but also how a medium, such as AVR, filters and affects representations of others during a mediated social interaction [8]. Note that these others can be either real people or virtual agents. Social presence is an important aspect in VEs due to its impact on social influence, which is how the presence of others modify individuals' attitudes, beliefs or behaviour [65]. Continual awareness of others in a VE is argued to be required for users to be able to flexibly adapt their behaviour regarding the other in social situations [19]. Spatial presence is concerned with social presence as well in the sense that spatial presence is hypothesized to increase with the existence of other individuals in the virtual environment and with the number of interactions between a user and another user or agent [19].

By including the therapist in the experience by means of AVR, emotions and experiences can be directly exchanged between client and therapist and allows for the therapist to immediately support the client while they are exploring an IVR experience. Regarding the utilization aspect of exploration, this support can for example be helping clients determine how certain situations make them feel and what makes these situations risky, explaining possible coping strategies and how to apply them, or role-playing a character relevant to the client to enhance a depicted social situation. For the navigation aspect of exploration, treatment providers can support clients by for example helping them locate objects relevant to elicit cue reactivity, such as a specific brand in a supermarket, or objects related to a specific coping strategy, such as a game for distraction.

Having more information on SUD treatment and the possibility for using IVR and AVR to support it, the next step is to determine what is necessary for communication for both the navigation and collaboration aspect of joint exploration to properly enable this support.

2.2 Navigation

Navigation is an ambiguous term amongst researchers as well [46], making it necessary to first establish its meaning within this research before continuing with focusing on navigating in a group and the accompanying communication.

One common way of defining navigation is for it to be the combination of wayfinding, a term originally introduced by Lynch [55], and motion. These two terms are roughly described as motion being the motoric element (*i.e.* movement) of navigation and wayfinding the cognitive element (*i.e.* tactical and strategic parts that guide movement) [11, 22, 60, 59]. As denoted by Darken and Peterson [22], navigation is rarely a primary task and tends to get in the way of what people actually want to do and that it is often the goal to make the execution of navigation tasks as transparent and trivial as possible, whilst not precluding the elements of exploration and discovery. To them, exploration and discovery can only take place if conditions of disoriented people, such as anxiousness and discomfort, can be avoided.

The process of navigation can be broken down into several components. The model of navigation as proposed by Darken and Peterson [22] is argued to be the most generally applicable. The model can be found in Figure 6. In short, the model works as follows, using the shoe shopping example they provided. In the model you start out with a goal (*e.g.* I need to find new shoes). Then you need to come up with a strategy to reach that goal (*e.g.* shoes can often be found at department stores, department stores are typically on the far points of the mall). What follows is the wayfinding/motion loop: you perceive the environment (*e.g.* look around, find a map), assess progress towards the goal (*e.g.* check the map), and determine how to guide your movement (*e.g.* see that you need to take a left). During this loop you can build a cognitive map of your environment (*e.g.* remember how to get back to the entrance). During motion or assessment you can determine to change the goal or strategy (*e.g.* I cannot buy shoes, lets buy lunch - the map was not helpful, I should ask someone knowledgeable). Note that during navigation this wayfinding/motion loop is repeated many times, as any new information observed in the perception phase is to be taken into account and new information is constantly provided in dynamic worlds. This includes new information that can be provided by others, either actively or passively. This social aspect of wayfinding is what we focus on next.

2.2.1 Social wayfinding

Dalton *et al.* [21] indicate that there is a substantial amount of research done on social wayfinding by means of verbal instructions and collaborative decision-making in context other than navigation, but a surprising lack of research on the social aspects of wayfinding.

They addressed social wayfinding in two ways: the nature of interaction between actors and the time frame in which the interaction takes place. They argue that the nature of interaction between actors, and as such the presented

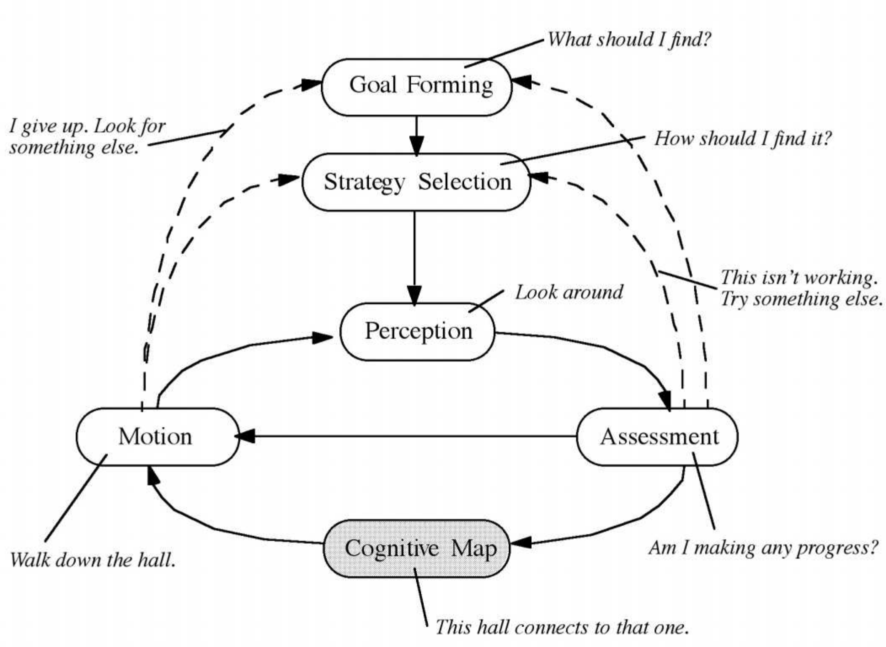


Figure 6: A model of navigation as proposed by Darken and Peterson [22]. Each bubble depicts a component of the navigation process, the arrows indicate the flow between the components, the grey bubble indicates that that component is optional during the process, and the dashed arrows indicate optional decisions.

wayfinding cues, can be either *strong* or *weak*. They explain that weak wayfinding cues steer the direction of an individual passively, for example going where the crowd is going or following a path worn into a grass field. On the other hand, strong wayfinding cues involves intentional communication between actors, whether digital or human, about location or route choices by means of verbal, graphical, or gestural information. They state that, the primary distinction between the types is the degree of intentionality: strong intentional cues can be seen as collaborative where people unintentionally being cues themselves for the weak cues can be seen as less collaborative.

They further expand the classification by taking into account the time frame between production and observing of a wayfinding cue, called *time modes*. In the *synchronous mode*, a wayfinding cue is produced and directly observed, whereas in the *asynchronous mode*, the produced wayfinding cue is observed at a later point in time. For example, asking directions from a local is synchronous and looking at a bought map is asynchronous. Combining the types of interaction and time modes, they classified four types of social wayfinding: synchronous strong, asynchronous strong, synchronous weak, and asynchronous weak, as is shown in Figure 7.

They use this information to build forth upon the taxonomy of aided and unaided wayfinding as presented by Wiener *et al.* [100], who argue that the type of wayfinding task, and cognitive strategies that can be applied to solve a task, are based on three aspects: (i) existence of external aid, (ii) having a specific destination, and (iii) availability of three different levels of knowledge. The knowledge levels are: having knowledge about the location of a specific goal (destination knowledge), knowledge about a specific path towards a goal (route knowledge), and knowledge of the lay of the land (survey knowledge).

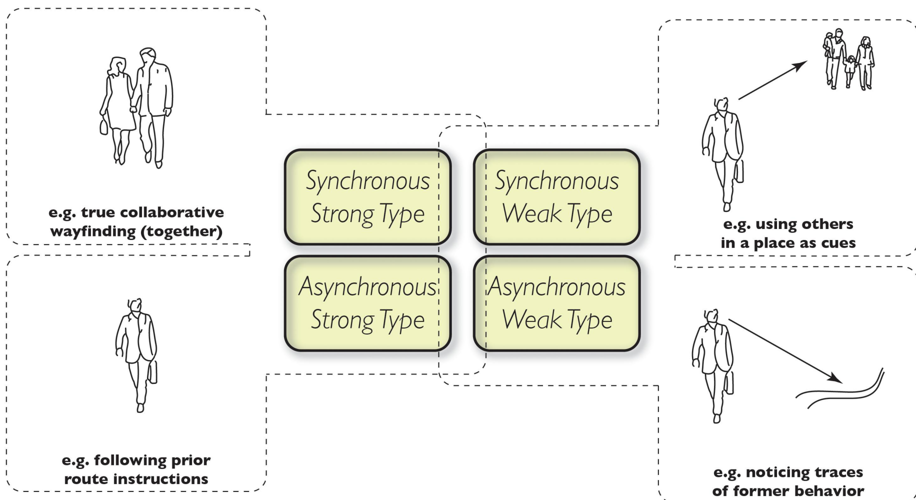


Figure 7: Four types of social wayfinding as proposed by Dalton *et al.*[21]. The yellow bubbles indicate the types and the dashed lines indicate the four quadrants and captures an example for each quadrant. The overlap, and lack thereof, of dashed lines are an indication that information of multiple types can be used simultaneously during wayfinding.

Dalton *et al.* [21] adds that there are three aspects that influence strong synchronous wayfinding in groups of two people, *i.e.* dyads, as well: (i) disparity of spatial knowledge, (ii) disparity of *sense-of-direction* and *cognitive styles*,[5] and (iii) disparity of character properties such as social status or personality. They explain that with a certain combination of these factors someone might assume a leadership role such as having superior spatial knowledge or having an higher social status within the group (*e.g.* boss and employee). Being more equal in these factors allows for greater mutual decision making.

For (i), Reilly *et al.* [72] investigated the coordination of the use of a digital map in pairs with varying spatial knowledge and found that a leader/follower style would emerge if the spatial knowledge varied and a more collaborative style emerges if both members of the dyad had similar spatial knowledge. For

[5]Cognitive styles as proposed by Aggarwal and Woolley [1]: psychological dimension that represents consistencies in how someone acquires and processes information.

(ii), He *et al.* [39] looked at people describing routes over a cellphone with the dyads having varying levels of *sense-of-direction*. They found that dyads that differed greatly in their sense-of-direction had a harder time completing the task and noted a difference in landmarks mentioned between the levels of *sense-of-direction* as well. For (iii), Dalton *et al.* [21] explains if one person in the dyad is of lower status, that person is less likely to question the wayfinding decisions of the higher status (*e.g.* child and parent or boss and employee) or a naturally dominant person might assume leadership despite possibly being less suitable for the wayfinding task.

Up until now we have mainly concerned ourselves with the understanding of navigation and social wayfinding in real life, however, in virtual environments such collaborative opportunities are not granted. Therefore we take a look at supporting collaborative work in VEs as well.

2.3 Collaboration in VEs

Elements that characterize collaborative work in VEs must be addressed in the design of the VE, as anything that is not accounted for will simply not exist in the VE. In addition to performing actions and receiving feedback through their individual interface, users also require information about other users' actions and activities through what is called *awareness* [30].

2.3.1 Awareness

As Fraser *et al.* [30] make clear, people sharing a virtual environment need to be aware of the activity of other users in order to help them to understand the evolution of the environment and to collaborative more efficiently with other users. The concept of being aware of the activity of others is referred to as *awareness*.

Awareness is supported by *awareness cues*, which are observable cues indicating the activity of others in the virtual environment. Cues regarding information about location and perspective can be provided by, for example, a virtual embodiment in combination with gaze direction and visualized field of vision [25, 71], a *what you see is what I see* (WYSIWIS [84]) window [25], or an observable miniature version of the active virtual environment [25]. Examples of cues for providing information on actions and communication are virtual pointers and hand gestures [70]. An important aspect related to awareness that should not be overlooked is the notion of social conventions of transitions, such as approach and departure in collaborative virtual spaces [13]. They suggest that collaborators should not violate normal social conventions of approach, for example abruptly entering or departing the collaborative space. As suggested by Piumsomboon *et al.* [70], people normally rely on their awareness of peripheral vision and spatial sound to sense the direction of approach. Having covered awareness and social conventions, we look further into the implications of navigation to support joint exploration in VEs.

2.3.2 Collaborative navigation

Specifically for the navigation aspect of joint exploration, Yang and Olson [105] determined four common characteristics or task requirements that contribute to collaborative spatial information-seeking activities, what they refer to as *collaborative navigation*. The four characteristics are: (i) users have to be able to remember or recognize the environment to some extent, (ii) each participant has an independently controlled view of the environment, (iii) there is a need for users to converge to a common visual or spatial location, and (iv) it is beneficial to understand others' perspectives. (i) refers to remembering spatial features of the environment such as overall shape or object distribution in order to use them, as they argue the importance of a shared mental model of the environment. The other three characteristics focus on the interplay between visuals of multiple users. For (ii) it is important to recognize that what they consider as *view* includes the view*point*. For example, a child sitting on their parent's shoulders has an independently controlled view as they can move their head independently and their viewpoint is being controlled by the movements of the parent. However, Yang and Olson would consider the child's view being controlled by the parent. (iii) and (iv) both relate to user perspectives, which can vary substantially and more easily in virtual scenarios compared to real life, and thus we will look into how certain combinations of perspectives affect collaborative navigation.

2.3.3 Perspectives

Tasks often require use of multiple representations of a virtual environment, each tailored to a different point of view (POV) and different sub-tasks [18]. For example, during social wayfinding, someone might look at a map with various environmental details to plan a route whilst another person looks around for corresponding landmarks to determine whether the planned route is feasible. These points of view can be divided into egocentric and exocentric perspectives. An *egocentric perspective* is the perspective as observed through one's eyes, *i.e.* a first person perspective, whereas an *exocentric perspective* is the perspective of a remote observer looking in on the environment, *i.e.* a third person perspective. For combinations of perspectives, a combination of two matching perspectives seems to support understanding between users [76, 105], as people might struggle to perform the mental spatial transformations required to understand the others perspective [105]. Yet, the combination of an egocentric perspective paired with an exocentric perspective, where the exocentric perspective has an overview position, lets the users take advantage of each other's viewpoints for strategical spatial tasks, such as wayfinding [76, 105, 107]. Additionally, a *what you see is what I see* (WYSIWIS) [84] approach seems to be the best way to correlate views. However, a WYSIWIS approach restricts the potential of collaborative work due to the lack of independent viewpoints [105].

We have built a theoretical rationale on supporting SUD treatment with AVR and the necessary communication for both the navigation and collaboration aspect of joint exploration. In Chapter 2.4 we focus on other works and practical implementations of systems relevant to this research, to get a better idea of the works and notable results of what has already been done in regard to the relevant topics of this research. Thereafter the findings as described in this chapter as a whole are discussed.

2.4 Related work

Works and notable results that are most relevant for us to describe are those that combine all three topics of AVR, joint exploration, and use for SUD treatment. Due to the limited number of works that combine all three topics, works that relate to two of these topics are considered as well. To not widen the scope of what is related too much, we limit ourselves to only include works that incorporate some sort of collaborative AVR system with a combination of the following aspects: having either co-located or spatially separated users, the collaborative space is either a VE or real life, and the work possibly includes the wayfinding aspect of exploration or SUD treatment.

We start out by describing works that relate to this research most, *i.e.* works with co-located AVR, VE as collaboration space, and either with or without inclusion of wayfinding or SUD treatment. Thereafter, we briefly summarize a non-exhaustive number of works with spatially separated AVR with either a VE or real life as collaboration space, and either with or without inclusion of wayfinding or SUD treatment that lie relatively close to this research to provide a rough idea of what else is there.

2.4.1 Works with co-located AVR, VE as collaboration space, and inclusion of wayfinding or SUD treatment

For the combination of the three aspects that relate the most to this research, we identified four works ([35], [54], [83], and [106]), one commercially available game (Black Hat Cooperative[6]) with a focus on navigation as well, and one commercially available AVR system which focuses on IVR therapy such as SUD treatment (CleVR[7]). In the systems of most works, non-immersed users get the possibility to observe the VE via a static monitor ([35], [54], [83], Black Hat Cooperative, and CleVR), but Grasset *et al.* [35] and Lee *et al.* [54] showed the possible use of AR and a projection mapping based approach respectively as well. In all approaches except for CleVR, the non-immersed user has an exocentric perspective on the environment with the IVR user having an egocentric perspective. CleVR shares the egocentric perspective of the IVR user with the non-immersed user in a WYSIWIS manner. Regarding the interaction possibilities, all aforementioned solutions allowed verbal communication between users, although Stafford *et al.* [83] prohibited the use of verbal communication in certain experiment conditions. Where the systems vary most from each other is in

[6] Black Hat Cooperative on Steam: `https://store.steampowered.com/app/503100/Black_Hat_Cooperative/` [visited 16-May-2021]

[7] CleVR by CleVR B.V.: `https://clevr.net/` [visited 20-December-2021]

how the system enables non-verbal communication, although Lee *et al.* [54] and Grasset *et al.* [35] only allowed verbal communication. Zenner *et al.* [106] used an IVR controller as flashlight to provide clues, Stafford *et al.* [83] captured hands in close proximity to a horizontal screen and displayed them as if coming out of the sky in the VE, in CleVR non-immersed users are allowed to control the emotions and conduct of virtual characters by choosing from a predefined list and speak via the characters using a voice-modulator, and Black Hat Cooperative had the additional option to send text messages that were shown directly on the HMD view.

Notable results are that Stafford *et al.* [83] determined that visual cue-based collaborative navigation was significantly more efficient than audio-only techniques and that natural hand gestures are more expressive and were superior to cursor-like visuals for task completion times. Grasset *et al.* [35] shows that monitor see-through AR, here referring to users that see objects on a monitor and can position their hands behind the monitor to interact with the objects, was significantly more efficient compared to tablet AR or HMD AR. Lee *et al.* [54] specifically looked into the feeling of presence by both users and determined that their system was able to elicit similar levels of presence and still have individual experiences in their cooperative game where two players controlled different aspects of a single character. In this game the IVR user moves the character in a maze and the non-immersed user performs necessary actions with the character using their hands. CleVR is used in various studies over the past ten years of its continuous development.[8] However, these studies seem not to focus specifically on the interaction between therapist and clients via the system, but rather used the system as a tool for therapy related research.

2.4.2 Works with co-located AVR, VE as collaboration space, but no inclusion of wayfinding or SUD treatment

Similar to the works that included a focus on navigation and SUD treatment, all works we found with this combination of aspects allowed verbal communication between users and used a practically static display or projection to visualize the VE for non-immersed users ([6], [37], [41], [68], and [89]). The other works we found with this combination of aspects used a dynamic projection mapping approach [75], three touch displays attached to the HMD [38], and next to a general floor projection, Gugenheimer *et al.* [37] also provided non-immersed users with a small display attached to an IVR controller as a more dynamic window into the VE as well. Regarding perspectives, Gugenheimer *et al.* [37] and Thoravi Kumaravel *et al.* [89] provided non-immersed users with an egocentric perspective as opposed to an exocentric one, while the remaining works provided an exocentric perspective.

Again, there was much variety in how interaction was provided to the users. Thoravi Kumaravel *et al.* provided non-immersed users a WYSIWIS view including a short timeline to make it possible to scroll back to a frame that was

[8]Use of CleVR in various media: `https://www.clevr.net/media.html` [visited 20-December-2021]

previously observed by the IVR user such that depth-corrected hand-drawn annotations could be provided in the VE. Peter *et al.* [68] tested various highlighting techniques to support guidance scenarios. Gugenheimer *et al.* [37] provided non-immersed users with an IVR controller with viewing window to interact with the VE and the IVR user, similar to Roo and Hachet [75]. In Gugenheimer *et al.*'s FaceDisplay [38], the non-immersed user could provide input using the touch displays attached to the IVR user's HMD or using a hand-tracking sensor which is also attached to the HMD. Lastly, in the work of Beimler *et al.* [6] an IVR tour guide could move IVR users to spots in a larger VE, that the IVR users could not reach when walking their dedicated action space themselves, by dragging IVR users using a top-down view of the VE on a large horizontal screen.

For these works, notable results are as follows. In the evaluation of Peter *et al.* [68] on various highlighting techniques they found that coloured outlines worked best for guiding the attention of the IVR user in combination with a virtual representation that provided the outline, a virtual drone in their case. They also note that the virtual representation of the non-immersed user being present positively affected the presence of the IVR users. Horst *et al.* [41] determined that 2D representations of users in a 3D environment already provides enough information for mutual visual communication. Gugenheimer *et al.* [37] suggests leveraging the asymmetric differences between AVR users as opposed to trying to level the experiences while the power of each user over the other remains the same. They add that high interpersonal dependency can slow down experiences while a lack of interpersonal dependency can separate experiences and incorporation of physical proximity in experiences can increase presence at the risk of breaking the IVR illusion with possibly incoherent audible or tactile information. In a later work, Gugenheimer *et al.* [38] found that having the non-immersed users interact with the IVR user using the touch displays and hand-tracking sensor attached to the HMD resulted in a high level of dominance and responsibility for the non-immersed users. Lastly, Thoravi Kumaravel *et al.* [89] found increased efficiency in a block assembly task using their approach and both Beimler *et al.* [6] and Roo and Hachet [75] indicated potential for their approaches but indicated to require a more formal evaluation.

2.4.3 Alternative variations of aspects

As the other variations of aspects sit further away from the scope of this research, we briefly summarize a non-exhaustive number of works that lie relatively close to this research to provide a rough idea of what else is there. Unfortunately, no works that include SUD treatment in these alternative variations of aspects were identified. We start with works that include AVR for spatially separated users, VE as collaboration space, and inclusion of wayfinding.

In the scope of the 2012 IEEE 3DUI Contest[9] challenge, researchers set out to create collaborative navigation systems for use in VEs, with Bacim *et al.*

[9] 3DUI 2012 Contest. In Proceedings of Symposium on 3D User Interfaces (3DUI). IEEE, 2012.

[4], [14], Nguyen *et al.* [61], and Wang *et al.* [96] presenting AVR approaches in which non-immersed users used static displays or projections to provide visual cues from an exocentric perspective to IVR users. Visual aids such as light sources [14, 61], arrows [61, 4], path markers with lines connecting them [96], or warning signs and blockades [4], were all deemed sufficiently useful for enhancing collaborative navigation in VEs.

In a similar combination of aspects, Horst *et al.* [42] and Thanyadit *et al.* [88] focused on immersive learning experiences, and more specifically on providing awareness cues to teachers. Horst *et al.* [42] looked at ways to make visually clear what a student is looking at and has already looked at and found a preference for having a small WYSIWIS window and visualization of the field of view (FOV) in the VE. Thanyadit *et al.* [88] recognized that such visualization techniques would result in visual clutter when shown for multiple students simultaneously and tried to algorithmically place representations of students in a more appealing way as opposed to showing their actual location in the VE.

Then there are works which focus on real life as the collaboration space between spatially separated users. With the focus on real life, IVR technology is used to provide a remote user some sort of 3D representation of the surroundings of the other user. To provide this representation, Kasahara and Rekimoto [47] and Lee *et al.* [53] used a 360 degree camera mounted to the non-immersed user and Kolkmeier *et al.* [49], Piumsomboon *et al.* [69], and Piumsomboon *et al.* [70] provided a virtual reconstruction of the collaboration space. All works except that of Kasahara and Rekimoto [47] provided the non-immersed users with AR glasses to allow them to observe visual information provided by the remote immersed user, where Kasahara and Rekimoto's solution only allowed verbal communication, which they determined to result in a lack of social presence provided by the remote IVR user. The other works show that provided awareness cues and gestural cues helped substantially in mutual understanding and overall collaboration [69, 53, 70], the importance of being able to create a common point of reference, which lacked in [53], virtual embodiments of the IVR user positively affect social presence experienced by the non-immersed user [49, 70], and view independence between users positively affects task performance [49].

Lastly, the works of Amores *et al.* [3], Gao *et al.* [32], Huang and Alem [43], and Tecchia *et al.* [86] are examples of works that specifically focus on supporting local users with hand gestures from remote users, who were only provided a 2D or single depth image frame of the real world environment, and the gestures were presented back to the local user through AR or a separate screen. All these studies presented some sort of preliminary results in which is made clear that gestural support does add to the collaborative experience, but they do not go as deep into other effects.

2.5 Discussion

In this chapter we set out to build an initial design rationale by provisionally answering SQ1 through SQ3 and looking at related work. Up until this point we have acquired information on SUD treatment, AVR, navigation, and collaboration in VEs as well as related work. Now, we discuss these findings in relation to answering SQ1-3 and determining the implications of related work as a whole. What the implications of specific findings are on the design of an intended solution are discussed in more detail in the next chapter, Chapter 3.

2.5.1 Implications SUD treatment on AVR design

For SQ1, we try to determine what the critical aspects are that an AVR solution needs to adhere to to make that solution suitable for use in SUD treatment. To get us closer towards answering SQ1 we looked at theory related to SUD, AVR, and IVR for SUD treatment. We identified an importance of IVR technology being able to provide ecologically valid environments and eliciting cue reactivity [51], which is helpful in approaches that focus on neuroplasticity that are used in SUD treatment. Note that ecologically valid environments refer to environments that correspond to their occurrence in everyday life and neuroplasticity refers to the brain's ability to learn or extinct conditioned responses, here being specifically reflexive responses to substance-related cues [64], such as craving or increase in heart rate [17], *i.e.* cue reactivity.

Whether a VE is perceived as ecologically valid and is able to elicit cue reactivity was found to be seemingly related to the notion of presence, but that relation appears to be relatively understudied for SUD treatment. For VRET, used for treatment of various phobias and anxiety disorders, spatial presence was deemed seemingly necessary for successful treatment outcomes [62], however, a causal relationship is yet to be determined [36]. For SUD treatment, a positive correlation between spatial presence and craving levels of smokers [26, 27] and a positive correlation between craving levels and perceived ecological validity [80] were found. Note that in [80], perceived ecological validity was considered a part of presence together with spatial presence, the feeling of being psychologically involved, and any adverse physiological reactions such as VR sickness. These are promising results for an apparent relation between presence, specifically spatial presence and perceived ecological validity, and SUD treatment outcomes, but it is clear that more research focused towards this relation is required. Despite the lack of concrete evidence in previous research as of yet, we still consider presence for clients using IVR to be a critical aspect for AVR SUD treatment, considering the explanation of why people respond realistically to IVR as explained by Slater [82].

Slater introduced the concepts of place illusion (PI) and plausibility illusion (Psi), where PI is similar to spatial presence and Psi is the extent to which a system can produce events that directly relate to someone. They argue that when both PI and Psi occur, people will respond realistically to the VE. It is important to understand that they argue that for neither PI or Psi to occur the VE needs to be realistic. For example, if the VE is a purple coloured forest where the animals can talk to you, PI and Psi can still occur resulting in realistic responses of the user.

When the VE is able to let PI and Psi occur, the VE should also be able to

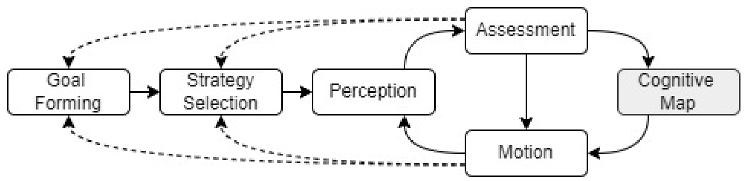


Figure 8: A concise version of the model of navigation as proposed by Darken and Peterson [22].

elicit cue reactivity, as cue reactivity is a realistic response. This means that cue reactivity can occur in VEs that are not necessarily ecologically valid as well. For example, in the purple coloured forest there are bottles hanging from the trees that verbally beg the user to drink them. This scenario might elicit cue reactivity, but is not ecologically valid or realistic, making that VE seemingly less suitable for SUD treatment. An ecologically valid scenario that might elicit cue reactivity is for example a football match and your friend offering a drink. As such we argue that the occurrence of PI and Psi are seemingly necessary for eliciting cue reactivity for SUD treatment, but only in VEs that are ecologically valid as well.

Presence being a relatively fuzzy concept, with an apparent role in IVR SUD treatment, and the argued connection with PI and Psi, we consider the following. We consider obtaining and maintaining presence, and the occurrence of PI and Psi for clients using IVR, critical aspects that an AVR solution should accommodate towards for the solution to be suitable for SUD treatment.

2.5.2 Communication for joint navigation

For SQ2, we aim to determine what is necessary for effective communication on navigation. To be able to answer SQ2, we looked at how navigation can be explained by means of wayfinding and motion, how the navigation process can be modeled as proposed by Darken and Peterson [22] and what the implications are of making navigation a social activity as explained by Dalton *et al.* [21]. What is not evident from this information is a clear idea of the content of the communication, which is what we are going to attempt to establish first using Darken and Peterson's model of the navigation process. Figure 8 shows a more concise version of the model for reference.

In their model, the navigation process starts by forming a goal and determining a strategy to reach that goal followed by a perception and motion loop to get to that goal whilst constantly assessing if progress towards the goal is made with the chosen strategy and cognitively mapping the environment. In either the motion or assessment step it can be determined that the process needs to start over by choosing a different goal or strategy. When navigating with multiple people, supposedly each person goes through the navigation process individually, but arguably every step of the process can be done collaboratively except for cognitively mapping the environment, as that happens individually in the brain. Of course, allowing for collaboration does not necessarily mean that collaborating individuals will have similar processes, as they can still have different goals, apply different strategies or have different ways of perceiving the environment to reach the same goal quicker together for example. Therefore, it seems beneficial to enable communication regarding the related *wayfinding* steps: establishing common goals and strategies, exchange what is perceived, discuss assessed progress, and motion.

The wayfinding task and potential strategies, and arguably the communication thereof, are in turn influenced by three aspects as argued by Wiener *et al.* [100]: (i) the existence of external aid, (ii) having a specific destination or (iii) the availability of either destination, route, or survey knowledge. Within navigation, motion is separate from wayfinding and therefore not part of the communication; motion is the result of said communication. Motion can still be collaborative if one helps the other in motion, for example by lending a hand to pass difficult terrain or being the driver of a car and the other being a passenger. One could say that discussing how to pass difficult terrain would require communication in the motion step. However, in the model this is represented by restarting the process with as goal to pass the difficult terrain, where communication occurs in the other non-motion steps. Dalton *et al.* [21] provides us an approach on *how* this information regarding wayfinding can be communicated.

Dalton *et al.* shows four types of social wayfinding that differ in the nature of interaction between actors, being either strong or weak, and the time frame in which the interaction takes place, being either synchronous or asynchronous. The type that is relevant for this research is the strong synchronous type, which they explain involves intentional communication between actors on location and strategy and where produced wayfinding cues are directly observed. This communication can be done by means of strong synchronous wayfinding cues: verbal, graphical, or gestural cue. On how strong synchronous wayfinding is influenced in dyads, *i.e.* groups of two people, they identify three aspects: (i) disparity of spatial knowledge, (ii) disparity of sense-of-direction and cognitive styles, and (iii) disparity of character properties such as social status or personality. Different variations of aspects and variations of aspects in specific contexts all results in different variations of influence.

Looking at these aspects in the context of SUD treatment, all three aspects highly depend on the therapist and client themselves as well as the content of an exercise and whether that exercise was done before. What we can argue for (i) is that a therapist wants full control over the VE and as such wants to have full survey knowledge of the VE and that the level of knowledge of the client depends on how familiar they are or have become with the VE or how much knowledge the therapist wants to share. For example for the latter, it might be important for the therapist to see what objects or locations a client can find for themselves before providing knowledge on what can be found where. For (ii), it is likely that the cognitive style is substantially different between therapists and people with MBID, but the therapist should know how to adapt themselves to facilitate specific cognitive styles that accompany MBID. Lastly for (iii), the power relation between therapist and client seems similar to that of a teacher and student, resulting in the therapist supposedly having authority over the client, which might result in less mutual decision making [21]. Similar to the disparity in cognitive style, we think a therapist should be able to balance this power relation in the best interest of the client.

Thus for effective communication on navigation for SQ2, people should be able to communicate about the goal, strategy, what they perceive, and assessed progress using strong synchronous wayfinding cues, which are verbal, graphical, and gestural cues. This communication is influenced by the existence of external aid, having a specific destination, available knowledge levels, disparity of spatial knowledge, disparity of sense-of-direction and cognitive styles, and disparity of character properties such as social status or personality, which all partly depend on the specific context of the navigation task.

2.5.3 Communication for collaboration in VEs

For SQ3 we aim to determine what is necessary for effective communication for collaboration in VEs. We noted the importance of supporting awareness in VEs by means of awareness cues [30] and taking into account social conventions of transitions [13], the contributions of certain characteristics or task requirements for collaborative tasks as presented by Yang and Olson [105], and the role of perspectives and dynamic between combinations of perspectives in collaborative tasks [18]. Notably, Yang and Olson characteristics or task requirements were set up with collaborative *spatial* information-seeking activities in mind for a fairly dated idea of data representation in 3D VEs. However, looking at the basis on which these characteristics were set up, they can arguably be generalized to joint exploration of VEs in general, including the utilization aspect of exploration.

The four common characteristics or task requirements as presented by Yang and Olson are: (i) users must be able to remember or recognize the environment to some extent, (ii) each participant has an independently controlled view of the environment, (iii) there is a need for users to converge to a common visual or spatial location, and (iv) it is beneficial to understand others' perspectives. For them (i) refers to remembering or recognizing spatial features, such as the overall shape of the VE or object distribution in a VE, as they argue the importance of a shared mental model of the environment for collaboration. Looking at joint exploration in general, this shared mental model is not limited to spatial features alone and can well include interactive features of a VE. For example, recognizing and remembering the effect of pressing a specific virtual button or how one can interact and make use of an agent. Regarding (ii), it was argued to be necessary to best exploit the potential of collaborative work in general, so there is no further need for generalization. (iii) was set up on the premise of finding something interesting and wanting others to see it and similar to the generalization of (i), finding something interesting can relate to an utilization aspect of the VE as well. Lastly, (iv) was establish to avoid large communication overhead that is the result of not understanding what another is looking at or how one's viewpoint relates to the other. Again, this is not relevant for spatial exploration alone and applies to the utilization aspect of exploration as well. However, (iv) is focused on the *objective* visual view and does not cover understanding of another's *subjective* world view, which can be relevant for exploration as well. For example, two people exploring a virtual art gallery might be able to establish how they are objectively looking at an art piece, but might want to understand the other's interpretation of the art piece as well. Enabling certain spatial wayfinding cues, such as the strong synchronous cues discussed earlier, might already be sufficient for communicating the subjective world view to a certain extent.

In similar fashion, the model of navigation as proposed by Darken and Peterson [22] seems arguably applicable for the concept of exploration as a whole, including the utilization aspect. Making this generalization allows us to break down the process of exploration in general with the following three benefits: (i) it helps us get a better understanding of the content of communication regarding joint exploration as a whole, (ii) it makes it easier to determine if and how the necessities for communication specific to navigation might be suitable for joint exploration in general as well, and (iii) it can be taken into account in the design process and generalization of results of an intended solution to enable joint exploration for SUD treatment, which are all three relevant for attempting to answer RQ1.

We can make the navigational model applicable to exploration as a whole by generalizing the two navigation specific concepts within the model: motion and the cognitive map. In the navigational model, motion is the navigation related action that is taken to try to get closer towards the defined goal and as such we can generalize it to *take action* or *act* for short. The cognitive map in the model is an understanding of the spatial features of the VE. When generalizing we want to include both the spatial features as well as any other features, such as the utilization features, and therefore we generalize it as *cognitive understanding of the VE*, *understanding VE* for short.

With our current understanding of the exploration process and communication possibilities, as a proof of concept, a more practical implementation of the generalized model regarding a therapist and client practicing coping strategies together can be described as shown in Figure 9. The implementation represents a summary of the multitude of cycles that happen throughout the whole exploration process as one cycle. For example, in the client's *perception step* they check if the VE contains anything SUD related and if that is not the case during assessment, the goal can temporarily shift to finding SUD related elements in the VE first before being able to apply coping strategies. Additionally, the *assessment step* describes assessment of how the VE affects the client as well as the assessment of the ability of applying coping strategies, which would normally occur in separate iterations of the process. Considerably the most important step in the cycle for SUD treatment is the *understanding VE step* for the client, as it directly relates to neuroplasticity, *i.e.* the ability of the brain to learn or extinct conditioned responses. Whether and how concepts that were set up in the context of navigation, such as strong synchronous wayfinding cues or quantification of spatial knowledge into three levels, apply to this specific implementation are discussed in the chapter on ideation, Chapter 3.

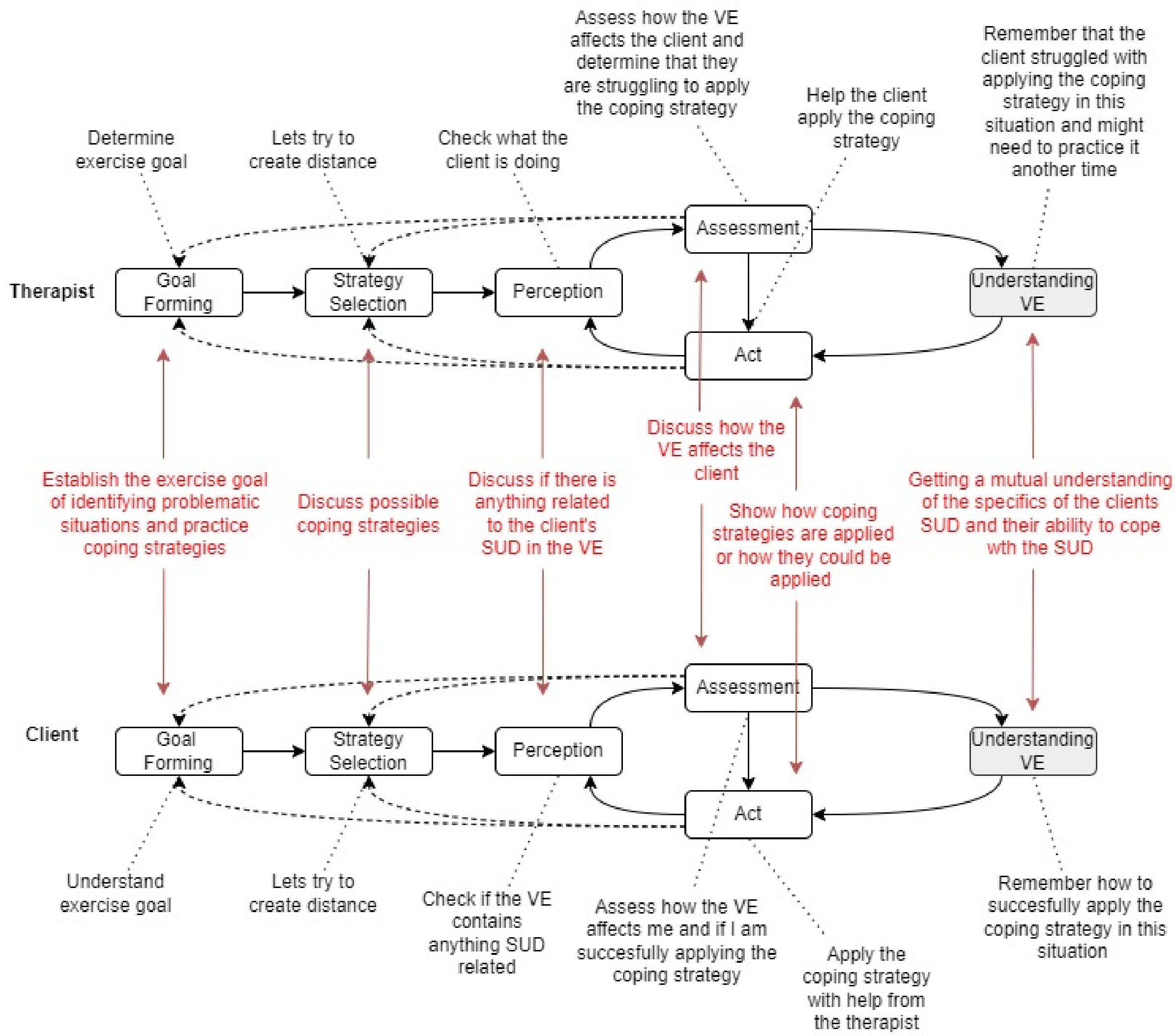


Figure 9: Implementation of two generalized exploration models, based on Darken and Peterson's [22] navigational model, in parallel representing a therapist and a client practicing coping strategies together. The top process represents that of the therapist and the bottom process that of the client, as indicated on the far left. The red arrows and text indicate collaborative opportunities regarding the implementation for the process steps they point towards. The black text with the dotted lines show an elaboration of each step regarding the implementation. This implementation represents a summary of the multitude of cycles that happen throughout the whole exploration process as one cycle.

2.5.4 Insights related work

Regarding related work, the combination of the AVR and exploration topics and accompanying variations in those topics, such as including a navigation aspect or having spatially separated users, resulted in a high variety of relevant works. These works provide a wide spectrum of ideas that can function as inspiration when designing an envisioned solution, but it makes it clear that the search for a more general solution, if there is any, is still ongoing as well. Regarding works that focus on AVR for SUD treatment, CleVR is seemingly the only established solution. Using the CleVR system, treatment providers can help clients explore scenarios. However, they seem to only be able to do this by role-playing characters relevant to the scenario or as themselves verbally without a visual representation in the VE. Representing the therapist as themselves by a character in the VE seems possible given the implementation, however, that usage is not discussed. Yet, they would seemingly still lack abilities that benefit collaboration and joint navigation, such as a separate perspective or gestures to help converge to a common visual or spatial location. The implications of other valuable insights works might have provided are discussed during ideation and prototype design in their respective chapters, Chapter 3 and Chapter 4.

Concluding, in this chapter provisional answers on SQ1-3 were obtained by discussing theory related to AVR for SUD treatment, joint navigation, and collaboration in VEs and an idea of how this information could be practically used was sketched by discussing related work. Following the hourglass analogy, we have narrowed the scope far enough to start discussing the practical implications of the information we obtained with regard to ideation of a potential solution in the middle section of the hourglass.

3 Ideation

In this chapter the ideation phase of the first UCD iteration is described. Within the first iteration, the aim is to create a baseline system that covers the basic needs for an AVR solution to support joint exploration for use in SUD treatment. We start by identifying stakeholders which is followed by the requirements engineering process in which we discuss the implications of the information obtained in Chapter 2 and how these implications translate to requirements. Thereafter we discuss the premises on which we based our brainstorm and the three most promising ideas that followed from the brainstorm. We choose one idea and conceptualize it further into what will become the first iteration prototype. A full description of the first iteration prototype and evaluation thereof is discussed in the next chapter, Chapter 4.

3.1 Stakeholders and requirements engineering

Before we can start requirements engineering, stakeholders must be identified, as they are an integral part of the requirements engineering process. Stakeholders are all individuals or organizations that are impacted in some way by the new product or who influence the requirements of this product [33]. However, as many details of a potential product are still unknown, such as who would further develop or maintain the product, we focus on the baseline stakeholders as defined in [79]: individuals or groups that are going to use the system in one way or another or have an extensive stake in the project as they are decision makers of some sort. We identified three baseline stakeholders, of which two fall within the scope of this research:

- **SUD clients with MBID (at Tactus)** are stakeholders as they are intended main end-users of the solution, which could be incorporated in their treatment. Clients are the users of IVR technology within the solution and will therefore have the role of IVR user (IVRU) in the context of system design.
- **Treatment providers (at Tactus)** are stakeholders as they are intended main end-users of the solution as well, as they will be using the system as part of treatment procedures and have the responsibility of aiding clients before, during, and after use of the solution. Treatment providers are the users of non-immersive technology to partake in the IVR experience and will therefore have the role of externally partaking user (EPU) in the context of system design.

Note Tactus as an organization or clinic is considered a baseline stakeholder as well, as they have an interest in the development of new technologies that could support or improve current treatment approaches and they could change their treatment plans based on the impact of the solution, which would require further development and maintenance that require an investment. However, changes to treatment plans and deploying a potential solution within the organization fall outside the scope of this research and as such, Tactus as an organization is not considered in the requirements engineering phase and no business-level requirements will be defined.

3.1.1 Eliciting requirements

Having identified the baseline stakeholders, we can start looking at the first step of the requirements engineering process as defined by [63]: eliciting requirements. Eliciting requirements consists of gathering initial information from stakeholders in order to be able to formulate requirements. There are various techniques to accomplish this information gathering that are focused directly on stakeholders, such as questionnaires, interviews, and workshops. However, there are also indirect techniques that include analyzing existing documentation and prototyping. In this research we opted for using solely the indirect approach of analyzing theory and related works, as presented in Chapter 2. We opted for this approach for two reasons: (i) we think that a concept which combines the themes of AVR for SUD treatment and both navigation and collaboration of joint exploration is best presented to end-users in the form of a tangible prototype for them to understand and reason about, as opposed to presenting an abstraction of this combination, and (ii) the discussed theory seems to provide sufficiently enough information to set up requirements for a baseline prototype for the purposes of (i), which can be further refined in a second UCD iteration. With hands-on experience, end-users might be better able to identify the strengths and weaknesses of the baseline prototype in relation to use for SUD treatment of people with MBID at Tactus.

Next are the modelling and analyzing requirements steps. In these steps, information gathered in the elicitation step is processed involving visualizing relations between requirements and classifying requirements to be either user requirements, functional or non-functional requirements [99]. Functional system requirements specify what must be implemented so user requirements can be fulfilled, and non-functional system requirements specify subjective operation attributes as opposed to behaviour. We do this by first discussing the implication of the information obtained in Chapter 2 after which we translate these implications into requirements.

AVR for SUD treatment

For SQ1, we determined that obtaining and maintaining presence and the occurrence of PI and Psi for clients using IVR are critical aspects that an AVR solution should accommodate towards making the solution suitable for SUD treatment. To make these aspects more tangible we identified four factors that play a role in obtaining and maintaining presence: (i) interface awareness [104], (ii) real-world consistency [104], (iii) social presence [40, 73], and (iv) VR sickness [97]. (i) Interface awareness regards the effortless interpretation of and interaction with the VE, which might be hindered due to unnatural, clumsy, or artifact-laden interfaces, which should thus be avoided. (ii) Real-world consistency is about providing information and interaction similar to the real world in order to retain presence, which is in line with providing ecologically valid experiences. (iii) The social presence of others, being either real or fake, and the ability to interact with their virtual embodiments can increase presence, which also relates to the occurrence of Psi. Lastly, (iv) VR sickness is caused by visually mediating self-motion but failing to mediate other organs that perceive self-motion. As such, an IVR locomotion technique needs to be chosen that addresses this factor.

Furthermore on SUD treatment, the treatment provider wants full control over the VE and within the scope of this research. That means full control over their actions in the VE, which includes the level of provided knowledge in the VE. Therefore the treatment provider must be able to determine the level of provided knowledge in the VE themselves and the solution must not be unpredictable in the amount or depth of information it produces in the VE. More specifically, this includes the information an embodiment of the treatment provider provides as well. For example, when they want to see how a client handles a situation by themselves. Additionally, one of the benefits of using IVR is that clients can practice in an environment that is ensured to be safe. To ensure that an environment is safe, the treatment provider needs to remain aware of their surroundings. Lastly, considering common traits of people with MBID, the way clients act and interact via a solution should be easy to understand and remember, as any cognitive effort spent on using the solution is arguably better spent on the SUD treatment itself.

Joint navigation

Information specific to navigation was argued to be applicable to the exploration process as a whole and as such, we consider any following requirements to be applicable to exploration as a whole. To be able to explore the VE, the solution enables execution of the full exploration process, based on the navigation process as presented by Darken and Peterson [22], for both the client and treatment provider. Regarding communication on exploration between users, the solution should enable establishing common goals and strategies, exchange what is perceived and discuss assessed progress, for which strong synchronous wayfinding cues, *i.e.* verbal, gestural, and graphical cues, can be used to exchange relevant information.

With regard to the aspects that were determined to influence joint navigation, such as disparity of spatial knowledge or power relations, we recognize these differences, however, these aspects do not seem to translate into implications on the solution. The most considerable influence would be that of the power relation between treatment provider and client, where the treatment provider has arguably a higher status than the client. Treatment providers wanting full control over the VE could arguably include having control over the client's motion or actions in the VE without their consent as well. However, the resulting forced lack in autonomy for the client might create an unfavourable disposition between them and the treatment provider and we see no benefit in having that level of control over the client with regard to the discussed SUD treatment approaches. Therefore we refrain from giving the treatment provider control over the motion or actions in the solution without the consent of the client. Giving consent can be as simple as a verbal acknowledgement, which is to be enabled via strong wayfinding cues, and thus we think that the solution does not need to require anything more for providing consent as of yet.

Collaboration in VEs

For collaboration in VEs, Yang and Olson [105] provide us with four necessary characteristics or task requirements. Of these four there are three that are requirements for the solution: (i) having an independently controlled view of the VE, (ii) being able to converge to a common visual or spatial location, and (iii) being able to understand each other's perspective are requirements for the solution. Characteristic (i) regarding having an independently controlled view of the VE includes the viewpoint as well, and as such the characteristic also implies enabling motion for both users. The fourth and remaining task requirement, (iv) users need to be able to remember or recognize the VE by some extent, is not a requirement of the solution, but rather a requirement of its users.

The characteristic of being able to understand each other's perspective is especially important for the treatment provider, as understanding the client's viewpoint provides understanding of what affects the client. For example, if a client indicates to crave a substance upon visually perceiving a specific part of the VE, the treatment provider can immediately understand what that specific part is. We mentioned before that client and treatment provider might not constantly collaborate, for example because the treatment provider wants the client to try something by themselves and only observe what the client will do. For such situations, it is not necessary for the client to understand the perspective of the treatment provider, but for the treatment provider it is always necessary to understand the perspective of the client, resulting in two separate requirements for the solution. Understanding perspectives, amongst other things, can be supported by means of awareness cues.

Awareness cues are observable cues indicating the activity of others in the VE, which include cues for providing information about the location, actions, or perspectives of the other. Both users being able to produce and observe relevant awareness cues when collaborating is therefore required by the solution. Closely related to awareness cues is being able to comply to the social conventions of transitions, as one can announce their approach or department using awareness cues as well. Being able to comply to the social conventions of transitions avoids potential anxiety in scenarios which might already by stressful for clients to experience, essentially hindering the treatment.

3.1.2 Resulting requirements

The requirements that resulted from the modeling and analyzing requirements phase are presented in Table 2. The requirements were categorized on type:

user requirement (UR), functional requirement (FR), and non-functional requirement (NFR). Additionally, per requirement, the MoSCoW priority and origin topics are shown: AVR for SUD treatment (AST), joint navigation (JN), and collaboration in VEs theory (CVE). Regarding the MoSCoW prioritization, no requirements with a *could* or *won't* priority were identified. Requirements that were deemed necessary for support of SUD treatment, but not essential to enabling joint exploration or obtaining and maintaining presence by the client, were considered a *should* priority. Lastly, the functional and non-functional requirements show the requirement identification (ID) number of the user requirement which they support.

3.2 Idea generation

For idea generation, we first looked into current interaction possibilities in a conventional IVR setup to identify impediments that need to be accommodated towards in an AVR solution for joint exploration. Then a brainstorm was held to come up with ideas on how to shape the interaction between users, keeping in mind these impediments and potential implementation of user requirements. These ideas were then analyzed to identify general design properties that appear to be crucial for the solution. From these ideas, three ideas with most potential were chosen and discussed in more detail. We chose one of the three ideas as a starting point for the envisioned solution and sketched a more finalized concept according to an interaction scenario regarding joint exploration. Before describing the resulting prototype, we briefly discuss aspects of the first iteration prototype that were changed based on informal usability tests during development. A description of the first iteration prototype, how it accommodates the requirements, and its evaluation are described in the next chapter, Chapter 4.

3.2.1 Interaction possibilities

Looking at currently conventional IVR systems, specifically those that rely on a separate dedicated computer to render the VE presented by an HMD, interaction for joint exploration is often already a possibility, although in a substantially limited form. For example, it is often possible for the non-immersed user to watch along the perspective of the IVR user via a monitor, exchange information verbally, and converge to a visual location by letting the non-immersed user turn the IVR user physically towards a direction of interest. Figure 10 shows our model of the interaction process between IVR and non-immersed users using such an aforementioned conventional system. Our model of the process is build up as follows.

Before people can interact, that interaction needs to be initiated, which we believe be done either via a proposal or abruptly. We assume the proposal is visual and helps people acknowledge each other first, for example by moving closer and making eye contact, such that an initiation is not unexpected. We think an abrupt initiation is necessary if a person fails to acknowledge the other,

Table 2: Requirements of an AVR solution to support joint exploration for SUD treatment for people with MBID.

ID	Type	Requirement	Priority	Origin	UR ID
1	UR	Both user can execute the full exploration process independently.	Must	JN	-
2		Both users can communicate to establish common goals and strategies, exchange what is perceived, and discuss assessed progress.		JN	-
3		The treatment provider can always understand and act on the perspective of the client.		CVE	-
4		The client can understand and act on the perspective of the treatment provider during joint exploration.		CVE	-
5		The client can obtain and maintain presence allowing for PI and Psi to occur.		AST	-
6		The client can explore the VE without a representation of the treatment provider present.	Should	AST	-
7		The treatment provider can ensure the safety and confidentiality of their real surroundings.		AST	-
8	FR	Both users have an independently controlled view of the VE.	Must	CVE	1
9		Both users can exchange strong synchronous wayfinding cues.		JN	2
10		Both users are able to converge to a common visual or spatial location.		CVE	2
11		The treatment provider can determine the amount and depth of information they provide in the VE.		AST	2
12		Both users can exchange awareness cues during joint exploration.		CVE	2,3,4
13		The treatment provider can represent themselves via an embodiment.		AST	5
14		The treatment provider can control when a representation of them is perceivable.	Should	AST	6
15	NFR	The solution allows users to respect social conventions of transitions.	Must	CVE	1
16		The solution is predictable in the amount and depth of information it produces in the VE.		AST	2
17		The way interaction for joint exploration is enabled for the client is easy to understand and remember.		AST	2
18		The solution avoids interface awareness for the client.		AST	5
19		The solution provides information and interaction similar to the real world.		AST	5
20		The locomotion technique of the client avoids resulting in VR sickness		AST	5
21		The treatment provider can remain aware of their real surroundings and act on them.	Should	AST	7

which is the case in the conventional setup, where the IVR user is closed off from the physical world. This initiation seems to be only possible either verbally or physically in the conventional setup, such as calling the other's name or a tap on the shoulder. These abrupt initiations can easily startle the other as they are unexpected due to the lack of a proposal. After initiation, we argue that the interaction can take place, again using verbal and physical cues only. For example, by telling the other to turn to the right or rotating them physically to the right. When there is no need or desire for interaction anymore, we believe the interaction can be either ended abruptly or wrapped up first before ending. Wrapping up the interaction happens for example by telling the other goodbye and performing a handshake. The wrap-up seems to be especially important for interaction in the conventional setup, as for the IVR user there is no way to confirm that an interaction is still ongoing or not. For example, when ending abruptly, a non-immersed user could just walk away and the IVR user would not know if the non-immersed user is taking longer to respond or has left. They can only know by taking off the HMD themselves.

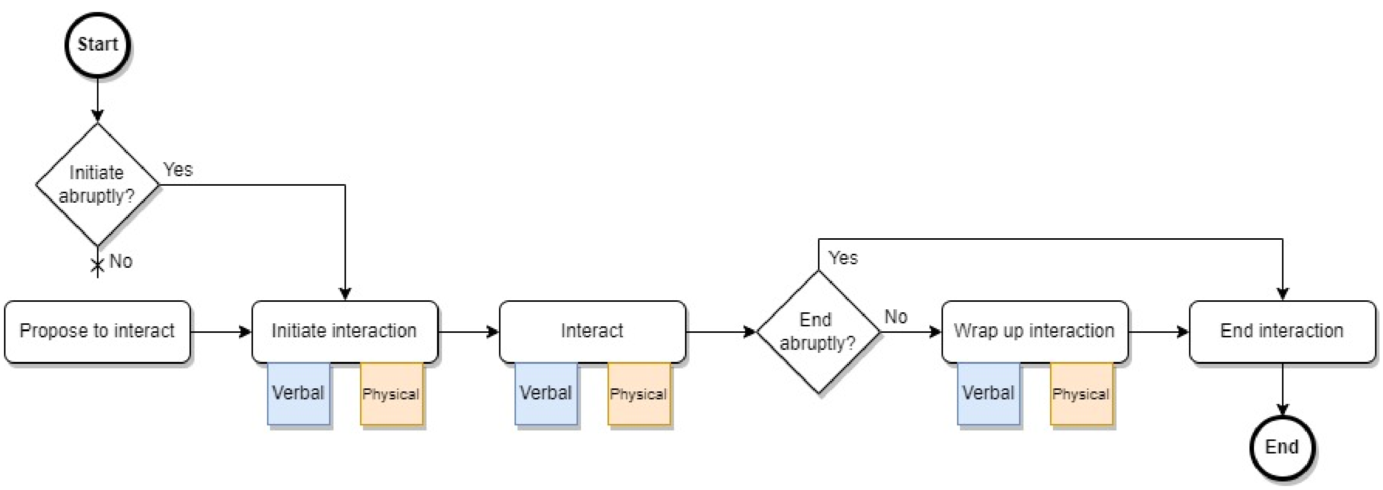


Figure 10: Process showing the interaction possibilities for interaction between users in a conventional IVR setup. The coloured squares below process states indicate which interaction types can be used. The line with a cross, below the first choice diamond, indicates that that flow option is unavailable using a conventional setup, resulting in a user only being able to choose the remaining option in the interaction process.

From this model we can determine that the visual aspect of interaction is severely lacking throughout the whole interaction process, not just the interaction itself. For example, a lack of visual awareness cues for the interaction proposal and wrap-up, or a wave gesture for the initiation. Next to limiting the interaction possibilities, it results in communication overhead as well, as for example demonstrative pronouns, *e.g.* this or those, cannot be used during verbal communication. Demonstrative pronouns are useful for converging to a common visual or spatial location. The lack of the visual aspect is an observation that will be taken into account during idea generation so we can ensure that each step of the interaction process is covered by a potential solution. This information was taken into account for the brainstorm that is discussed next.

3.2.2 Brainstorm results

We held a brainstorm to generate ideas on how to shape the interaction process in the VE. The resulting 29 ideas were analyzed to identify general design properties that appear to be crucial to consider for shaping this interaction. Sketches of the 29 ideas are shown in Figure 11 and a summary of the identified design properties is shown in Table 3. Instead of discussing all 29 ideas separately, we picked three ideas with the most potential for AVR interaction for SUD treatment for people with MBID in our opinion. A full overview of the identified properties for each idea are provided in Appendix A. The identified design properties are used in the discussion of these three ideas to better define and compare them when placing them in context.

Three ideas that stood out from the rest were those called: *Clone*, *Fairy flower*, and *Magic remote*. The original sketch of the ideas is shown in Figure 12. All three ideas try to capitalize on the possibilities of IVR and how it can go beyond what would be considered realistic for the benefit of the interaction, which was one of the benefits of using IVR for SUD treatment. We briefly discuss the core mechanics of each idea and their implications for joint exploration for SUD treatment after which we select one idea that we deem to have the most potential for the envisioned baseline solution to conceptualize further. Note that regarding the design properties we identified amongst all generated ideas, all three ideas are initiated by a single action, are not environment dependent, and have a humanoid EPU avatar representation which can move freely.

Clone

For the Clone idea, interaction is initiated by letting an avatar with the appearance of the IVRU emerge from their body. This clone of the IVRU can be controlled by the EPU, allowing the IVRU to perceive themselves from a different perspective and enabling a new form of self-dialogue, which can also help perceived self-efficacy. For example, the EPU could mimic problematic behaviour in a relevant scenario and show how to cope with that scenario thereafter as well to help the IVRU recognize their problematic behaviour and perceive themselves as competent in coping with the situation through effort. What stands out regarding the identified design properties is that we consider the interaction as a whole unfamiliar. While interaction with the other as a human being might be familiar, it is unclear what the implications are of interacting with what is supposedly oneself and what would be a understandable way of proposing and initiating interaction from the perspective of both the IVRU and EPU.

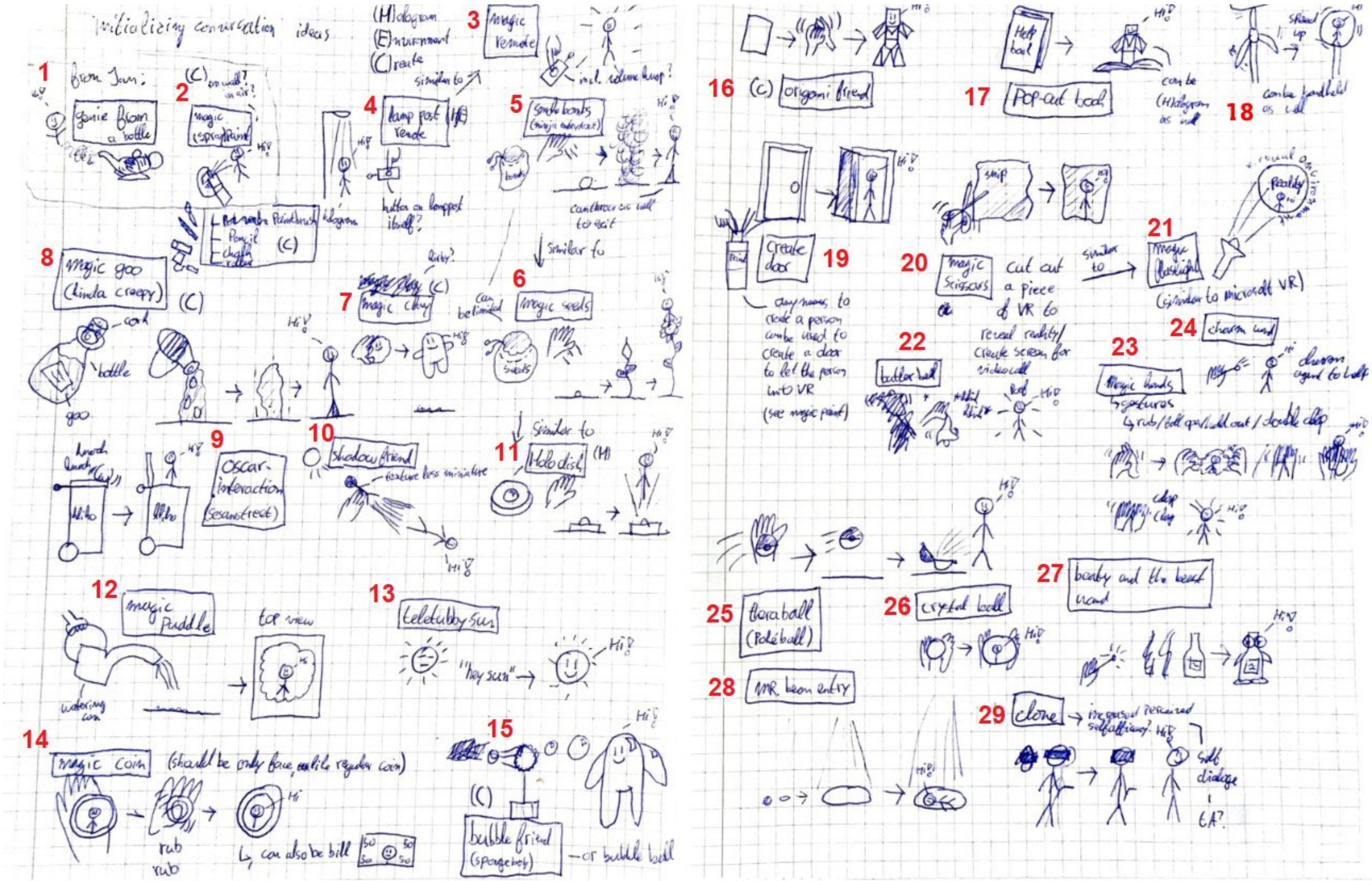


Figure 11: 29 sketches for ideas on entering and interacting in the VE as externally partaking users. Each interaction is titled as follows: genie from a lamp (1), magic (spray)paint (2), magic remote (3), lamp post remote (4), smoke bombs (5), magic seeds (6), magic clay (7), magic goo (8), Oscar-interaction (9), shadow friend (10), holo disk (11), magic puddle (12), teletubby sun (13), magic coin (14), bubble friend (15), origami friend (16), pop-out book (17), spinning fan hologram (18), create door (19), magic scissors (20), magic flashlight (21), butler bell (22), magic from hands (23), charm wand (24), therapist ball (25), crystal ball (26), Beauty-and-the-Beast wand (27), Mr. Bean entry (28), and clone (29).

Table 3: Identified interaction design properties

Property	**Description**	**Identified variations**	**Examples (identified variation)**
Initiation action	Action performed to get a representation of the EPU into the VE.	Perform a single specific action / shape a substance into a 3D representation / create a surface for a 2D representation	Rubbing a genie lamp (single action), forming a face out of clay (shape 3D), put magic paint on a wall (create 2D)
Environmental dependence	The spatial design of the VE is used to initiate and mediate the interaction.	Yes / no	A lamppost projecting a hologram (yes), create a shadow person by standing in the sun (yes)
Location of initiation	Which of the users choose the location of initiation.	Chosen by IVRU / chosen by EPU / relative to object	Planting a magic flower from which a fairy emerges (chosen by IVRU), EPU avatar parachuting down (chosen by EPU), lamppost projecting a hologram (relative to object)
EPU avatar appearance	The visual appearance of the EPU's avatar.	Humanoid / non-humanoid	Animate tree or stuffed animal (non-humanoid)
EPU avatar size	The visual size of the appearance of the representation of the EPU's avatar.	Human-sized / smaller / larger / surface dependent	A small fairy or the face on the side of a coin (smaller than human-sized), god-like from the sky (larger than human-sized)
EPU avatar freedom of movement	The extent to which the EPU avatar can move around freely in the VE.	Move freely / move within limits / static location	A dog on a leash (move within limits), displayed on a advertisement billboard (static location)
Requires carrying an object	The interaction requires the IVRU to always carry a specific virtual object with them to initiate or mediate the interaction.	Yes / no	Virtual smartphone (yes), genie lamp (yes)
Leaves a trace	Concluding the interaction results in a visual trace that indicates that the interaction took place.	Yes / no	A flower from which a fairy emerged (yes)
Limited number of uses	The interaction can only be initiated by the IVRU a limited number of times.	Yes / no	Limited smartphone battery (yes), limited magic flower seeds (yes)
Familiarity interaction	The extent to which the initiation or interaction is familiar for both users.	Familiar / unfamiliar	Smartphone (familiar), animating objects with a wand (unfamiliar)
Familiarity EPU avatar	The extent to which the EPU avatar is familiar for the IVRU.	Familiar / unfamiliar	EPU avatar looks like themselves (familiar), EPU avatar is a dog (unfamiliar)
Initiation deniability	Whether the initiation of interaction is deniable by both users.	Deniable / not deniable	Smartphone can ignored or hung up (deniable), EPU avatar parachuting down (not deniable)

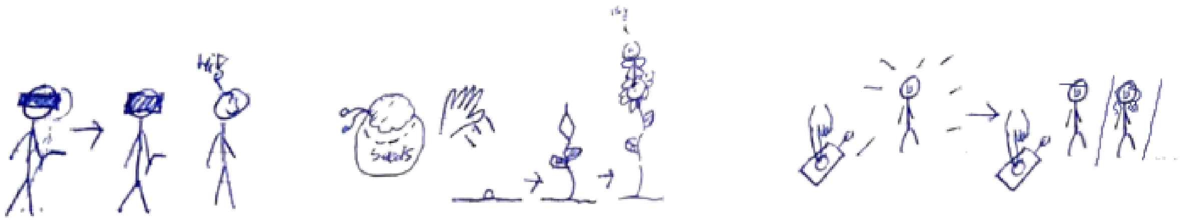


Figure 12: Sketches of the three ideas resulting from the brainstorm that have most potential. From left to right: Clone, Fairy flower, and Magic remote. In Clone, a clone with the appearance of the IVRU emerges from their body. In Fairy flower, a flower seed is planted from which a small fairy emerges. In Magic remote, the EPU uses a remote to make the EPU avatar appear who's appearance they can change via the remote.

Fairy flower
For the Fairy flower idea, the IVRU has seeds that when put on the ground quickly grow into a relatively large flower from which a small fairy flies up that is controlled by the EPU. After an interaction has finished, the fairy vanishes but the flower remains to remember the interaction by. The fairy's small size and their ability to fly allows for more flexibility with regard to movement in the VE for the EPU. For example, this allows the fairy to remain in the field of view (FOV) of the IVRU during an interaction so visual cues are always perceivable for the IVRU. The remaining flower can act as a mnemonic of the interaction such that when a similar flower is encountered a next time it might make it easier to remember and apply the things learned during previous sessions. This concept can be extended to real life as well, where clients can be given a flower upon leave of absence to help them recognize and remember their progress. We consider planting a seed a familiar interaction for most, however, being a magic seed, people might not recognize that it can be planted on any ground surface. The growth of the flower presents a clear gradual transition from initiating the interaction to the interaction itself, such that the IVRU knows what to expect and can prepare themselves for the interaction. If limiting the number of interaction possibilities is desired, this can easily be incorporated by limiting the number of seeds available to the IVRU.

Magic remote
The Magic remote allows the IVRU to bring forth the EPU avatar and using the remote they can determine the EPU avatar's appearance using the various buttons on the remote, allowing them to choose with who they are going to interact. Allowing the IVRU to choose themselves how the EPU looks will always ensure that the EPU is represented as someone they are comfortable with or of who they respect their opinion, such as the treatment provider, their parent, or a friend, which can also be different given specific situations. Of course, for each

individual depicted, there might be a certain expectation of behaviour during the interaction which the treatment provider might not be able to meet, due to a lack of knowledge of those individuals depicted or acting skills.

For our first iteration prototype we decided on using the Fairy flower idea as a starting point for our design. While the Clone and Magic remote seem to provide interesting opportunities regarding the EPU avatar appearance for SUD treatment, they also seem to be a step ahead of the interaction we are trying to establish for the first iteration baseline prototype. They seem to be a step ahead in the sense that they could be an interesting addition to an interaction we are still trying to establish. Looking at the interaction properties other than the EPU avatar appearance, the interactions of all three ideas are fairly similar in that they are initiated by a single action, are not environment dependent, and result in a humanoid 3D avatar that can move freely on their own.

These similarities indicate that our intuition for a baseline interaction went out to ideas with those properties. Considering that the fairy has the most utility in movement by means of flight compared to the other two ideas, it brings forth a relatively unfamiliar interaction as well, as it becomes harder to determine what to expect from their movement and might be harder to follow compared to the human-like avatars of the other two ideas. The implications of possibly having to carry magic seeds, limiting that number of seeds, and leaving a trace are yet to be determined once the interaction during conceptualization and testing during development is refined. Now that we have a direction for shaping the interaction, we can look into a fitting AVR setup to support that interaction.

3.2.3 AVR setup

The conventional setup as presented in Chapter 3.2.1 is in essence already a minimalist AVR setup, as the monitor provides the EPU a look into the VE, allowing for non-technologically mediated interaction between the EPU and IVRU regarding the VE using verbal or physical cues. After a brief brainstorm session on possible hardware implementations, we came to a similar conclusion as to shaping the interaction: using specialized hardware or techniques could be an interesting addition to an interaction we are still trying to establish. For example, interacting via hand tracking might allow for more elaborate gestures, but might limit the EPU in how they can move while gesturing, as we have seen in the work of Stafford *et al.* [83]. Therefore, for the first iteration prototype, we decided upon using the monitor, mouse, and keyboard already present in the conventional setup to enable interaction between users.

The EPU is required to use the conventional setup to interact, and is therefore bound to a relatively static location with respect to the IVRU. This might result in a discrepancy between the EPU avatar and their voice if not addressed. Incoherent audible was indicated by Gugenheimer *et al.* [37] to possibly break the IVR illusion. Therefore, we see benefit in adding a microphone for the EPU and sound dampening headphones for the IVRU so the EPU's voice can be perceived as if it originates from the EPU avatar in the VE instead of having a discrepancy between the location of the voice and avatar when not mediated.

This interaction can later be adjusted using different or more specialized hardware if deemed necessary during user tests. A sketch of this AVR setup is shown in Figure 13. Having determined a design direction for the interaction and AVR setup, we can further conceptualize the first iteration prototype by sketching out an interaction scenario regarding SUD treatment. Using this interaction scenario we can also ensure that all steps of our interaction model are covered.

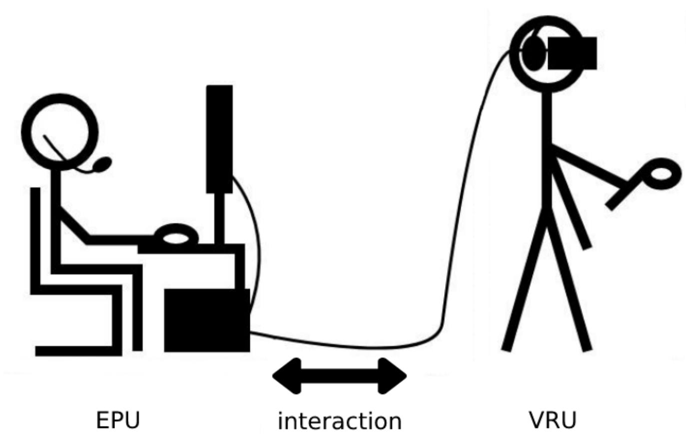


Figure 13: Sketch of the envisioned AVR setup for the first iteration prototype. On the left is the EPU using a monitor, mouse, keyboard, and microphone to interact with the IVRU on the left who uses an HMD, IVR controller, and sound dampening headphones.

3.3 Interaction scenario

In the following scenario description, a treatment provider and client are using the envisioned first iteration prototype to enable interaction for a SUD treatment exercise related to alcohol. The scenario assumes the client is introduced to the concept of IVR and knows how to use the technology beforehand. In the exercise, the client must go to a bus stop to get on the bus, but there are various alcohol-related elements along the way. The goal for the client is to practice with the 6D self-control techniques (*i.e.* Table 1) and thus gain insight on their behaviour and how to cope with that behaviour. The goal for the treatment provider is to gain insight on the actions and reactions of the client and help the client if anything is unclear or intervene when deemed necessary. The scenario description is supported by the visuals shown in Figure 14 and is a summary of the full scenario found in Appendix B.

Before the exercise, the treatment provider prepares the VE to make it suitable for the exercise and provides instructions to the client. Upon entering the VE, the client does not see a bus stop in their immediate vicinity (Figure 14-1) and rather than wandering they decide to ask the treatment provider for help by planting a magic seed to summon a fairy. The treatment provider lets the fairy point towards the bus stop and tells the client to go there verbally (Figure 14-2). The client indicates that that is all and the fairy may disappear.

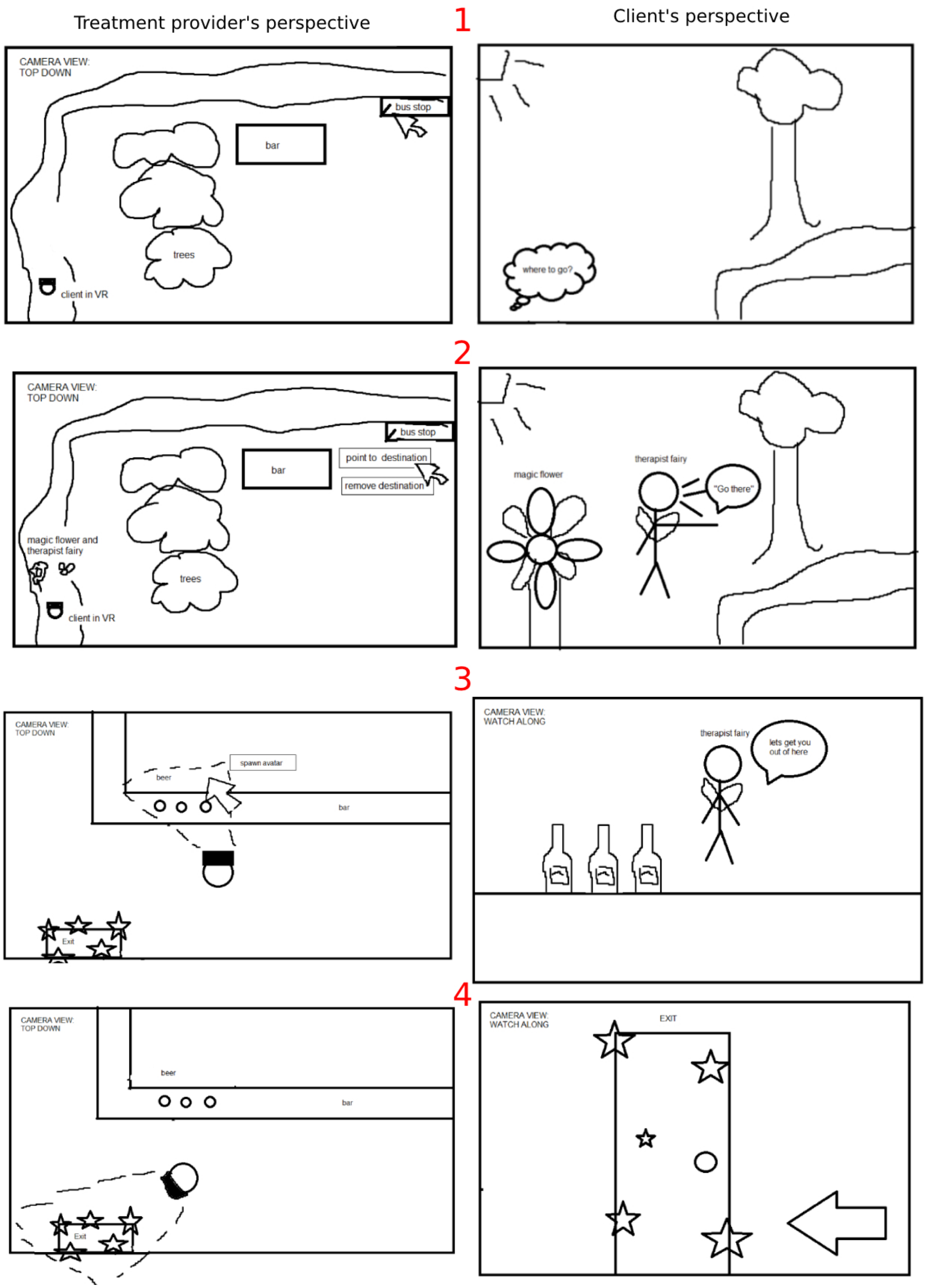


Figure 14: Four-step summary of the interaction scenario as presented in Appendix B. On the left the perspective of the treatment provider is shown and on the right that of the client. Step 1 shows a client in need of help to reach a bus stop, step 2 shows a fairy pointing the way, step 3 the client in the bar they encountered on the way and the treatment provider entering the VE themselves, and step 4 shows the fairy using graphical arrows and highlights to direct the client to the exit.

The client walks towards the given direction and they cannot resist the urge to enter the bar they encounter along the way. The treatment provider tries to hint that taking distance might be the better option here by highlighting the bar's exit. They check whether the client can see the highlight by visualizing their FOV and watching along their point of view (POV). Upon determining that the client does not notice the highlight the treatment provider decides to intervene by spawning a fairy themselves and telling the client to leave (Figure 14-3). In addition, the treatment provider shows the client that they need to turn around by showing a large arrow in their FOV. The client turns around and sees the sparkling exit (Figure 14-4) and decides to leave the bar and thereafter the bus stop is quickly found, finishing the exercise.

This interaction scenario shows all steps of the interaction model we proposed, including how the visual aspect in these steps can be covered as well. The IVRU initiates the interaction regularly by planting a magic seed, or the EPU initiates abruptly by simply making the fairy appear. The virtual embodiments of both EPU and IVRU can, for example, wave to initiate the interaction visually. The interaction between users was visually supported by gestures of the fairy and graphical sparkles. The interaction can be wrapped up visually, using a wave again for example, or abruptly by making the fairy simply disappear again, both ending the interaction.

Regarding the requirements we established, this interaction scenario shows a concept solution that covers all functional requirements, although some to a limited extent. Both users have independently controlled views (R8), can exchange strong cues (R9), can converge to a common location (R10), and can exchange awareness cues (R12) and the treatment provider can determine depth of information themselves (R11), has an embodiment (R13) and can control when that embodiment is perceivable (R14). As for the non-functional requirements, their subjectivity makes it impossible to reflect upon with such a scenario without introducing researcher bias. What can be said about the non-functional requirements is that the scenario describes interaction that goes beyond interaction similar to real life, conflicting R19. Having a theoretical proof of concept by means of the interaction scenario, we started development to see how well this concept translates to practice. Before we provide a description of the resulting first iteration prototype, we describe changes made to the provided concept during development.

3.4 Conceptual changes first iteration prototype

Based on informal usability tests during development, parts of the interaction changed to improve the interaction. We determined that enabling interaction between users using AVR is already a substantial design challenge on its own and that, similar to the Clone and Magic remote, the specifics of the Fairy flower aspect could be an interesting addition to an established interaction. The changes that this brought forth is that the EPU is represented by an avatar with a regular human appearance instead of a fairy and that only the EPU has control over where and when their avatar appears in the VE. This allows us to determine whether interaction with the human-like avatar is suitable for our purposes before going into more practical details such as proposal notifications, how to handle deniability of interactions, and 3D flight controls of the fairy without establishing their value in a user evaluation first.

Moreover, we identified that with a fully independent view of the VE for the EPU and the IVRU being able to move about the VE using teleportation locomotion, the spatial movements of the IVRU were hard to follow for the EPU, suggesting the need for EPU view controls that make it easier to follow the spatial movement of the IVRU. Following spatial movement is especially important considering that interactions were identified to consist of a singular longer interaction that covers multiple locations, in contrast to multiple shorter interactions on each location individually. The latter has implications for when to leave a trace of the interaction as well, as a trace that was left at the start of the interaction, such as the fairy flower, could already be out of sight when the interaction finishes, suggesting that traces should be left at the end of the interaction. Overall, the identified design properties are initiated by a single action, not being environment dependent, and result in a humanoid 3D avatar that can move freely on their own, while the covered functional requirements remain the same.

Concluding from our ideation process, our current concept is an interaction where the EPU uses a desktop to interact with the IVRU via an avatar with a regular human appearance. The avatar can be placed or withdrawn from the VE at any time independent actions available to the IVRU. During development we found that following the quick spatial movements by the IVRU due the their teleportation locomotion were hard to follow, suggesting a need for movement controls for the EPU that make it easy to follow the IVRU in the VE.

4 First iteration prototype

Based on the interaction concept and requirements established in the previous chapter, a first iteration prototype solution was created for enabling interaction for joint exploration for SUD treatment of people with MBID. This first iteration prototype is to be improved in a second UCD iteration using feedback from user evaluations. This chapter provides a full description of the first iteration prototype and how it accommodates the requirements first, followed by a description of the user evaluation and its results, and thereafter these results and their implications on design are discussed.

4.1 AVR setup components

Compared to the general description provided in the previous chapter, the only aspect of the AVR setup that has changed for the first iteration prototype is that the use of a keyboard is omitted. A more detailed visualization of the main and sub-components of the first iteration prototype is provided in Figure 15. The AVR setup consists of three main components: a desktop PC, desktop peripherals, and an HTC Vive set. On the desktop PC, the main application runs which handles in and output of both users such that they can both access a shared VE and interact with each other in that VE. The EPU uses the desktop peripherals, which consist of a monitor, computer mouse, and microphone, to use the application. They observe the VE and user interface (UI) via the monitor, perform actions in the VE using the mouse, and send verbal cues to the VE using the microphone. The IVRU uses the HTC Vive set, which consists of a pair of base stations, an HMD, a controller, and a pair of headphones. The base stations are not used directly by the IVRU, but are used by the HTC Vive to help the HMD and controllers track their exact locations better. In order for the main application to be able to process the IVR in and output, SteamVR[10] is used on the desktop PC. Using SteamVR, the base stations, HMD, and controller provide IVR positioning and input data such that the corresponding view and audio of the VE can be provided to the HMD and headphones respectively as well as actions resulting from controller input.

To enable interaction for the EPU, we opted for using mouse input only instead of using both the mouse and keyboard that are included in the conventional IVR setup. This was done to keep input complexity minimal considering the target user, namely treatment providers, for which computer proficiency is not a certainty. We believe that most desktop users know their way around mouse input, however, not everyone might be as adept in using the mouse and keyboard simultaneously and switching between mouse and keyboard input was deemed awkward, and also unnecessary, during development as all interaction could well be implemented using mouse input only. Using only the mouse, users must remember, or get used to, fewer input possibilities when using the application. Having this non-immersive setup should allow the EPU to remain aware

[10]SteamVR on Steam: `https://store.steampowered.com/app/250820/SteamVR/` [Accessed 19-April-2022].

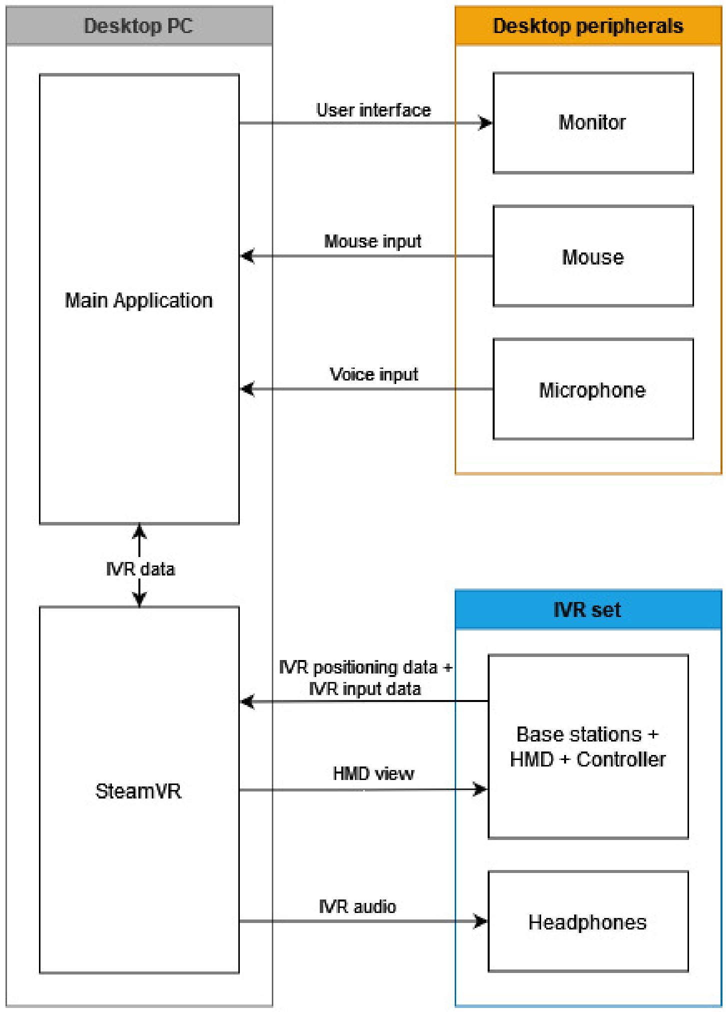


Figure 15: Main and sub-components of the first prototype and how they interact with each other. At the core lies the desktop PC, marked grey, which runs the main application and handles in and output for both users. The desktop peripherals, marked orange, are used by the EPU and the HTC Vive set, marked blue, is used by the IVRU. The black arrows between the components indicate the flow and what type of data is transferred between those components.

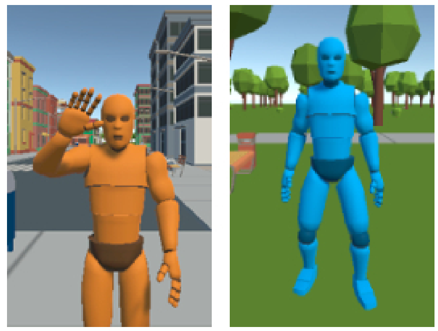


Figure 16: Examples of avatar appearance of both users. On the left, the orange avatar of the EPU performing a wave. On the right, the blue IVRU avatar in an idle pose.

of their real surroundings during use of the solution (R21). Understanding the components the AVR setup consists of and how these components are used by the EPU and IVRU, the interactions these components enable in the VE are described next.

4.2 Enabled interactions

In the following paragraphs is described how both users are represented in the VE by means avatars and how interaction is enabled for both users separately. Throughout these descriptions, we clarify how they relate to the established functional and non-functional requirements. How the solution's accommodation of the FRs and NFRs relate to the user requirements are discussed based on user evaluation results later in this chapter.

4.2.1 User avatars

In line with the need to enable awareness cues between users (R12) and the need for the EPU to be able to represent themselves via an embodiment (R13), both users are provided full-bodied human-like avatars according to the conceptualized idea. Because the focus of the prototype lies on enabling interaction and not on presentation, the avatars were given a neutral appearance similar to a test dummy. The avatar for the IVRU is coloured blue and the avatar for the EPU is coloured orange. These complementing colours should help the avatars stand out in the VE to be easily distinguishable. How these avatars behave based on user input are described for each interaction separately. Examples of the appearance of both avatars are provided in Figure 16.

We opted for full-body human-like avatars because they are easily recognizable as representing a human being and it makes it easier to determine what to expect from the avatar's behaviour. This allows for interaction that is similar to real life (R19), which might be easier to understand for the client as well (R17). The same could be achieved with using more photo realistic approaches, such as using real-time 3D image capture as in [41]. However, while perhaps being a more realistic representation of the user, a photo realistic representation does not necessarily add to the goal of providing awareness of each other in the VE. Such a connection to real life might hinder spatial presence from occurring and it adds unwanted system complexity for this stage of the research.

4.2.2 IVR

IVR technology provides the IVRU the ability to perceive the VE visually and audibly based on the position of the HMD (R8,10). The IVRU was provided a room-scale IVR play area in which they can walk around freely. With the play area being substantially smaller than the VE, the IVRU can also use teleportation locomotion to move the relative position of the play area inside the VE (R10). For teleportation, the IVRU uses the IVR controller to select a destination by pointing a beam on a specific location in their nearby vicinity. This beam is visible for the EPU as well, to have an indication of which direction the IVRU is going.

Teleportation locomotion was chosen for its ability to avoid VR sickness [29, 50] (R20). Teleportation does come with a potential decrease in presence [29, 50] and a decrease in spatial orientation [29], however, the avoidance of VR sickness has a high priority as it may cause discontinuation of the IVR experience. When looking around and teleporting in the VE, the IVRU avatar's position adjusts according to the position of the HMD relative to the VE. Additionally, the avatar's full body rotation adjust according to the yaw rotation of the HMD and the avatar's head adjusts according to the pitch rotation of the HMD. As such, the avatar provides information on where the IVRU is as well as their head pose and eye gaze for the EPU to perceive. The latter two are important communication cues, especially for the focus of attention [71], hence they have been made perceivable.

For the exchange of strong synchronous cues (R9), the IVRU has access to verbal and gestural cues. Verbal cues can be provided by the IVRU simply speaking out loud as the EPU is in the same room and is not wearing headphones. For gestural cues, the IVRU can move the IVR controller they are holding and the avatar's hand, arm, and shoulder will move accordingly using a predictive algorithm. Additionally, the IVRU can use the controller to grab and hold various objects in the VE.

Note that the IVRU is not presented any additional interfaces to avoid interface awareness (R18) and that their way of interacting should be relatively easy to remember (R17). How the VE is perceived and how gestural and verbal communication are enabled by the solution are similar to real life and as such should come naturally without the need to remember anything (R17,19). The teleportation and grab actions are considered relatively memorable as well, as they are activated by a single button press each, with the addition of positioning the controller to aim for the destination for teleportation (R17). Having only actions based on body movement and single button presses, the solution should be predictable in the amount and depth it produces in the VE as well (R16). By utilizing the IVR technology, we can provide the IVRU a solid basis for interaction regarding joint exploration with relative ease. Providing the necessary interaction for the EPU is less conspicuous.

4.2.3 EPU interaction dashboard

To enable interaction in the VE for the EPU, the EPU is provided a dashboard on the monitor that they make use of by using the mouse. The dashboard consists of six main parts as shown in Figure 17: the main interaction window, camera controls for the main interaction window, environment actions, EPU avatar and EPU avatar actions, IVRU awareness cue options, and the on-screen manual.

Figure 17: User interface of the dashboard of the first cycle prototype. The numbers denote the main parts of the dashboard: 1. main interaction window, 2. camera controls for main interaction window, 3. awareness cues of the client, 4. EPU avatar actions, 5. environment actions, and 6. on-screen manual.

To make the dashboard as a whole more intuitive for use, related buttons are grouped and each group has its own distinctly coloured and named window around it to easily identify which aspects of the application the buttons relate to when searching for a specific action or browsing through the available actions. For further clarification, each button presents a brief description together with an emoji-like graphic to indicate what the button is used for. Note that all text on the dashboard is in Dutch due to the intended target group of the user evaluation being Dutch. Translations are provided in text where necessary. The following paragraphs will describe the functionality of the dashboard in more detail.

4.2.4 Main interaction window and camera controls

Through the main interaction window, the EPU can view the VE from their own perspective (R8) and interact with the VE. In Figure 17, the window shows a part of a virtual city and the blue avatar representing the IVRU on which the main camera is currently focused. The main camera can be used in three different control modes, for which control hints are shown on the main window, and which automatically deals with environmental elements that block its view.

Main window camera and camera control menu
For the main window camera, we focused on controls that would make it easier to follow the IVRU avatar. Instead of controlling the position and rotation of the camera manually, the camera positions itself towards the position and rotation of a point of focus (POF). The position and rotation of this POF can be controlled in three different ways, each having a different degree of following the IVRU avatar: *fully follow*, *follow position and manual rotation*, and *manual positioning and rotation*. The EPU can switch between modes at any time while using the dashboard by pressing their corresponding buttons in the blue camera control menu below the main window, denoted by a *2* in Figure 17.

In the *fully follow* mode, the POF follows the position and rotation of the IVRU HMD, such that the camera looks towards the same direction as the IVRU. This mode can be used to automatically keep track of the IVRU avatar if there is no need for interaction. In the *follow position and manual rotation* mode, the EPU can control the rotation of the POF themselves while still following the spatial movements of the IVRU. This allows the EPU to observe the nearby environment independently themselves while not losing the IVRU out of sight. In the *manual positioning and rotation* mode, the POF's location and rotation are fully independent of the IVRU and can be entirely controlled by the EPU.

In all three camera modes, the camera can be moved along a predetermined vertical curved dolly track, allowing for the EPU to change their perspective relative to the POF, to either have more focus on the POF or the surrounding VE. A visualization of the different perspectives along the dolly track is shown in Figure 18. The top end of the track provides a high top-down perspective of the VE, changing to a giant's perspective further down, to an over-the-shoulder perspective closely behind the POF near the bottom end of the track. In the *fully follow* and *follow position and manual rotation* mode, when at the bottom end of the track, the camera switches to a WYSIWIS perspective of the IVRU's HMD. Having this control over their perspective allows the EPU to switch their perspective to best fit the specific exploration task, such as an overview perspective for a spatial exploration task or a WYSIWIS perspective to better understand the IVRU as was argued by Yang and Olson [105].

When controlling the camera position manually, the IVRU can still move around themselves and can potentially be lost out of sight by the EPU. When switching back to a mode that follows the spatial position of the IVRU, instead of instantly snapping to the position of the IVRU, the camera first moves to the high top-down perspective first and thereafter gradually moves to the location of the IVRU. Having this overview combined with the gradual movements should help the EPU get an understanding of where the IVRU is, or has gone to, in the VE relative to the position of where the EPU was before switching modes.

We noticed in our internal user tests that a direct coupling of position and rotation between the POF and IVRU HMD can become confusing and possibly nauseating quickly for the EPU. This is because of the quick spatial movements due to teleportation and quick or jittery head movements due to naturally using the HMD. To avoid this confusion and nausea we applied substantial damping to the coupling between POF and HMD to smooth out the jittery and instant movements, resulting in gradual transitions for large changes in position or rotation and negate acting on small changes. Yet, even with the damping, performing actions accurately while following the IVRU was determined to be a challenge. Therefore, whenever the EPU selects an action to perform, the camera mode automatically switches to Manual position and rotation such that the EPU can perform the action on their own pace without being dependent on the movements of the IVRU.

Mouse controls and control hints

When manually controlling the main camera, the EPU can change their perspective using the scroll wheel, rotate around the point of focus by holding the right mouse button and moving the mouse horizontally, and move the POF's position using edge-scrolling. Here, edge-scrolling refers to moving the camera view towards the direction of the window edge or corner the cursor is located. To make the available camera controls for a selected control mode clear to the EPU, the main window shows control hints as shown in Figure 19: small illustrations in the bottom right corner for rotation and changing perspective and bars with arrows around the window for edge-scrolling. Showing these control hints should make it easier for the EPU to learn and remember what ways of controlling the camera are available in each camera mode.

Handling obstructions

We noticed that when following changing spatial position of the IVRU, the VE could obstruct the IVRU by accident. Hence, large environmental elements, such as trees, buildings, or wall are made invisible for the EPU when the camera is near those elements. The elements made invisible for the EPU are still visible for the IVRU. Figure 20 shows an example of the obstruction removal in action. This feature ensures that the EPU can perceive the IVRU independent of their location and rotation.

Regarding the requirements, having the camera control modes and being able to change their perspective allows the EPU to independently control their view of the VE (R8) and helps the EPU converge to a visual or spatial location (R10),

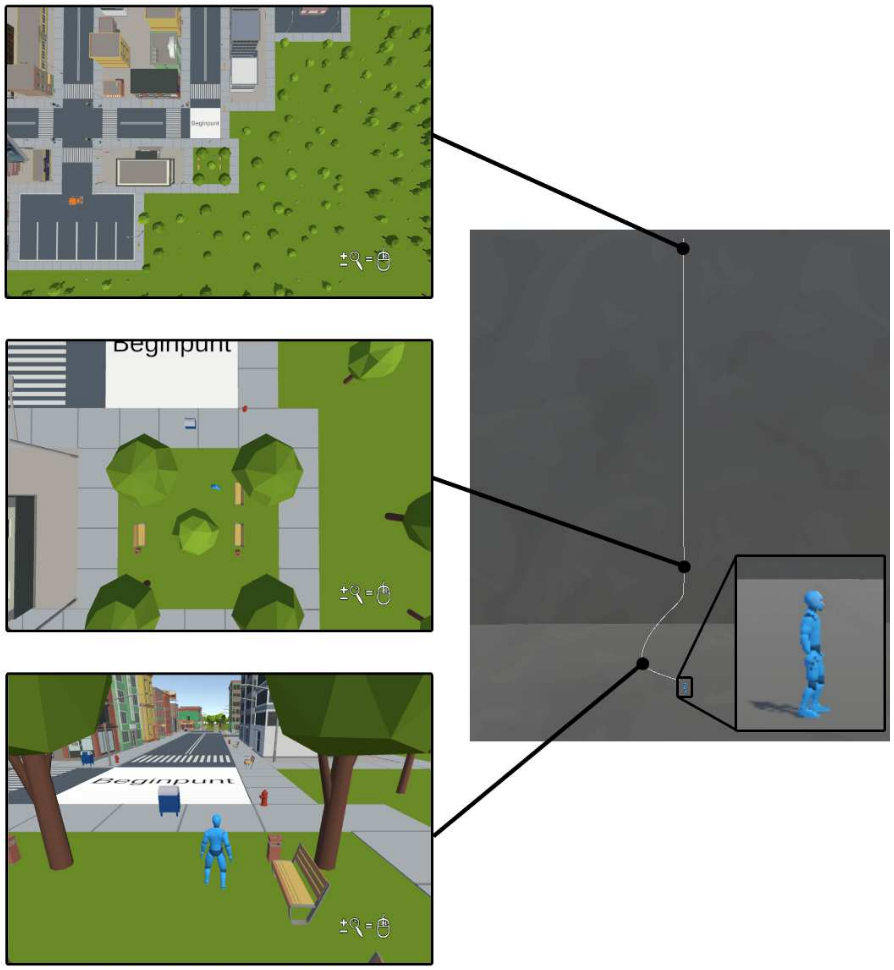


Figure 18: On the right, a visualisation of the dolly track, as a white line, relative to the point of focus, which is the blue figure in this example. On the left, snapshots of the main camera window perspective and the lines point to where these perspectives occur on the dolly track.

especially when following the movements of the IVRU. Having these varying perspectives, we noticed that the awareness cues produced by the IVRU avatar on head pose and eye gaze might not always be visible for the EPU, making it harder to converge to a common visual location. Repeatedly switching between the available exocentric and WYSIWIS perspectives of the main camera can become tedious and possibly confusing as well. Hence, we provided the EPU access to additional awareness cues about the view of the IVRU.

Figure 19: Main camera window showing all camera control hints: small illustrations for rotation using the right mouse button and changing perspective using the scroll wheel in the bottom right corner and bars with arrows around the window for edge-scrolling.

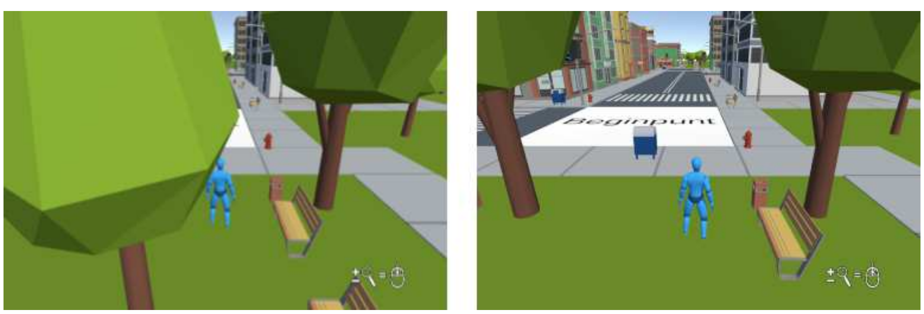


Figure 20: On the left, the IVRU avatar is visible, but a tree blocks a large portion of the main interaction window. On the right, the camera is moved slightly closer, making the tree invisible such that it does not block the camera's view anymore.

4.2.5 Additional awareness cues IVRU for EPU

Next to the IVRU's avatar and teleportation visualization, the EPU is provided three additional awareness cues on the IVRU's view: WYSIWIS window, viewing direction beam, and field of view frustum. Figure 21 provides a visualization of these cues. These cues were chosen as view-in-view, and volume based approaches for view visualization were preferred by participants in [42], and [71] showed benefits of the FOV and viewing direction visualization for awareness. The additional cues can be toggled on or off using a dedicated menu on the dashboard, denoted by *3* on the dashboard in Figure 17.

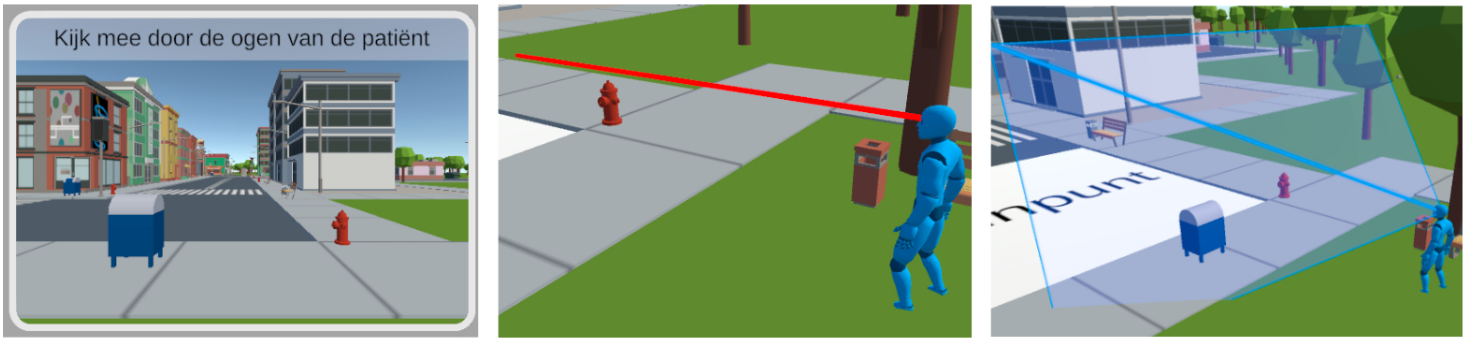


Figure 21: Optional additional awareness cues of the IVRU view. From left to right, the WYSIWIS window, viewing direction beam, and field of view frustum.

The WYSIWIS window resides in the bottom left of the dashboard and provides the most direct information on what the IVRU is currently seeing from their POV. Having the additional window might make it difficult to focus on one window with the other being a distraction, however, we did not experience this as a problem during development. The viewing direction is indicated by a long bright red beam. The sizable length of the indicator makes it possible to determine objects that are looked at more specifically, for example when there are multiple objects close to each other, the beam is likely to pierce the specific object that is focused on. The field of view is indicated by a blue frustum shaped volume. The volume is brightly outlined to clearly separate the volume from its surroundings. In contrast to the WYSIWIS window, the viewing direction beam and FOV frustum were visualized in the main interaction window. However, we noticed that they could become relatively intrusive in certain situations. Hence, the WYSIWIS window is still turned on by default, whilst the viewing direction beam and FOV frustum are now turned off.

Now that we have clear how the the VE and the awareness cues produced by the IVRU are perceived by the EPU, we look at the actions available to the EPU. The EPU has two types of actions available: actions that they perform via their avatar, called *avatar actions*, and actions that affect the environment that do not require their avatar, called *environment actions*. We start by describing the avatar actions.

4.2.6 EPU avatar actions

The buttons for the EPU avatar actions are located in the orange menu located in the top left corner of the dashboard, denoted by a *4* in Figure 17. In order to use the EPU avatar actions, the avatar needs to be present in the VE, which they are not by default, as we determined that the EPU should be able to control when their representation is perceivable in the VE (R14). Therefore, to use the avatar actions, the avatar needs to be placed in the VE first.

The EPU can place their avatar in the VE by clicking and dragging a picture of the avatar from the menu to a location of choice in the main interaction window. Upon selecting a location, a large cylinder fades in on that location and when that cylinder is fully opaque it fades away again revealing the EPU avatar. The cylinder fading produces a sound effect similar to a sliding door and a *ding* sound when being fully opaque. This process is visualized in Figure 22. The EPU avatar is removed by pressing the same button used for placement again, reversing the placement process without the *ding* sound, and leaving an empty spot where the avatar was. The cylinder appearing accompanied by the sound announces the avatar's arrival and department, such that the avatar does not appear suddenly in the vicinity of the IVRU, which is relevant for taking into account the social conventions of transitions (R15).

Figure 22: Placement process of the EPU avatar. From left to right, a location in the VE is selected by dragging a picture of the EPU avatar onto the main interaction window, a cylinder fades in until its fully opaque, the EPU avatar becomes visible in its place after the cylinder fades away.

Once the avatar is in the VE, it always positions itself towards the IVRU, automatically making eye contact with the IVRU while idle. This positioning and eye contact makes the avatar always look engaged with the IVRU, which is relevant for awareness (R12), but can help with Psi occurring for the IVRU as well, as it makes the actions of the avatar in the VE directly relate to the IVRU. Having their view and avatar separated allows the EPU to remain visually engaged for the IVRU while maintaining the benefits of having access to an exocentric and independent perspective of the VE. The other way around, the IVRU will likely appear visually engaged for the EPU, as they are still perceivable in the same room in real life whilst possibly separated in the VE.

The avatar's locomotion as well as verbal and gestural communication cues are enabled for the EPU after they have placed the avatar in the VE.

EPU avatar locomotion

The EPU avatar can move around in the VE using one of two modes: manual walking and following the IVRU avatar. Manual walking involves clicking a destination in the VE and the avatar will navigate itself towards that position. When following the IVRU the avatar will navigate itself towards the location of the IVRU and stop in their near vicinity. This process repeats once the IVRU teleports to a different location. During development, the avatar sped up to a running pace when needing to cover large distances to keep up with the possible speed of teleportation. However, we determined that running is likely to come over as anxious, which could make the IVRU unintentionally anxious as well. Hence, avatar locomotion is always done at a walking pace. In the same essence, the avatar always keeps a distance of at least one and half meters from the IVRU at all times to avoid accidentally invading the IVRU's personal space. When the IVRU covers large distances quickly, the EPU can recall the avatar and place it again when the IVRU reached their destination. When walking, the avatar's head and body are positioned towards the walking direction to make clear which direction they are going instead of towards the IVRU.

Verbal and gestural interactions

Using the avatar, the EPU can provide verbal and gestural information regarding the exchange of strong synchronous wayfinding cues (R9). Any microphone input provided by the EPU is played back in the VE from the avatar's mouth in 3D. Having stereo headphones, the IVRU will perceive these verbal cues from the direction the EPU avatar is located and its volume is adjusted based on the distance between the avatars of both users. The volume's lower limit allows the IVRU to always be perceivable while their avatar is in the VE independent of distance. The noise dampening headphones of the IVRU make it so that the real voice of the EPU is barely perceivable. While this does limit the connection with the real world to increase the IVRU's immersion, it makes it only possible to exchange verbal cues via the EPU avatar. Whether the microphone input is relayed is indicated by a microphone icon in the top right corner of the orange avatar action menu as well.

While the microphone is ideal for processing verbal cues, the mouse is less ideal for exchange of elaborate gestures. We chose to provide the EPU with a static set of predetermined gestures that we consider to be sufficient for enabling basic gestural communication regarding joint exploration for SUD treatment. The gestures are: pointing, waving, and applauding. We chose this approach for its simplicity to understand and to use while more complicated approaches, such as moving the arm manually similar to the IVRU, might not add to the interaction whilst likely harder to use given the 2D input of the mouse. Examples of the three gestures are shown in Figure 23.

The avatar is able to point towards specific objects or directions selected by the EPU and can be used to, for example, provide spatial knowledge or converge to a common visual or spatial location. The avatar can point or keep pointing while moving. When pointing, the avatar's head looks towards the object or direction pointed at instead of the IVRU or walking direction. The avatar's wave gesture can be used to greet the IVRU when initiating an interaction. The applause gesture, with corresponding sound effect, can be used to tell the IVRU that they are doing great, which is one of the 6Ds, named *applause* in the original Dutch variant.

Figure 23: From left to right, examples of the EPU avatar pointing, waving, and giving applause.

4.2.7 Environment actions

Next to the avatar actions, the EPU can also use actions that affect the environment for which the avatar is not required and of which the buttons are located in the green coloured menu below the avatar actions, denoted by a *5* in Figure 17. These action can be used to provide the IVRU with information without directly engaging with them or to provide supporting information while being engaged using the EPU avatar. These actions cover the graphical aspect of strong synchronous cues in three forms: object outlining, attention grabber, and temporary landmark. Examples of the three graphical cues are shown in Figure 24.

The outline and attention grabber were introduced based on the findings of Peter *et al.* [68], who found that coloured outlines worked best for guiding the attention of the IVR user. Objects selected using the outline action will have a bright green coloured outline that is visible through other objects, but is not perceivable when outside the view of the IVRU. Selecting an object through the attention grabber outlines an object and in addition makes a bright red arrow appear in the FOV of the IVRU that points towards the location of that object. This red arrow guides the attention of the IVRU towards objects even when they are outside the IVRU's FOV. Such an arrow was deemed intuitive and easy to learn by Wallgrün *et al.* [95], who compared approaches for attention guidance of IVR users. Both outline and attention grabber were implemented in accordance to the concept sketched in the interaction scenario and can for example be used to highlight objects of importance, provide wayfinding knowledge, or help

Figure 24: From left to right, from top to bottom, multiple object outlined using the outline action, a temporary landmark placed in the VE, the attention grabber highlighting an object behind a tree in the field of view of the IVRU, and the attention grabber indicating that the IVRU's attention should be redirected to the left.

converge to a common visual location. Lastly, the EPU can create temporary landmarks. The temporary landmark is a large flag pole with a checkered flag on top and due to its size, it should be visible in the VE from any height or distance. The temporary landmark was deemed useful for creating a navigational point of reference where existing landmarks would not suffice.

The dashboard providing this number of interaction possibilities might be overwhelming for inexperienced users. Hence the dashboard itself provides additional information on how to use it.

4.2.8 On-screen manual

In order to help the EPU find their way around the dashboard either before or during use, the dashboard provides access to a digital manual, denoted by the number *6* in Figure 17, and provides additional information on the purpose of each button when hovering over the button for a short time. The manual can be opened at any time during use of the application without hindering application usage by the IVRU.

Having described the full first iteration prototype, the next step is to evaluate this first iteration prototype in a formal evaluation to see how the theoretical design rationale holds for practical use.

4.3 Evaluation first iteration prototype

The next step in the process is to perform an evaluation with the realised prototype. The purpose of this evaluation is twofold: (i) gain insight on how to improve the interaction experience of the first iteration prototype within the context of SUD treatment for people with MBID and (ii) determine the implications of the design rationale we established by answering SQ1-3 and looking at related work when placed in the SUD treatment context.

We established separate evaluation questions to support the process of determining these implications. The evaluation questions are as presented in Table 4. Note that EQ4 covers the experience of presence by both users and not just the IVRU. While the experience of presence for the EPU was not identified as a critical aspect for SUD treatment, it is relevant with regards to the treatment provider remaining aware of their real surroundings (R21) and AVR experiences outside the context of SUD treatment. What follows is a description of the evaluation that was done to attempt to answer these evaluation questions.

This evaluation was approved by the ethical committee of Electrical Engineering Mathematics and Computer Science of the University of Twente (RP 2021-182) and the Wetenschappelijke Commissie at Tactus Verslavingszorg.

Table 4: First iteration evaluation questions

ID	Evaluation question
EQ1	How accommodates the solution critical aspects that an AVR solution needs to adhere to to make that solution suitable for use in SUD treatment?
EQ2	How are interactions regarding joint navigation experienced by the users?
EQ3	How are interactions regarding collaboration experienced by the users?
EQ4	How were the notions of presence and social presence experienced by the users?

4.3.1 Participants

Within the system there are two roles to fulfill, the EPU and the IVRU. Within the context of the use case, it is important that participants fulfilling the EPU role are treatment providers experienced with treating people with MBID for SUD. For the most complete picture of the interaction in context of the use case, the IVRU role should be fulfilled by said clients, however, it was decided to postpone the involvement of this vulnerable group until a more refined interaction experience was reached. Instead, the role of IVRU is fulfilled by a non-vulnerable research assistant that will behave similar to a client.

The involved treatment providers should be able to speculate about how real clients would experience the IVRU role, given the the interaction experience provided in the evaluation and a brief introduction on the IVR experience itself. Treatment providers are recruited on a voluntary basis at Tactus Verslavingszorg. No participants were promised any kind of compensation, and no compensation was given at the end of the experiment.

4.3.2 Approach

An attempt at answering the evaluation questions is made using the following approach. To let participating treatment providers experience the solution, an exercise is played out together with the research assistant. This exercise should contain elements of both the navigation aspect of exploration as well as the utilization aspect in the context of SUD treatment. An exercise that involves both aspects simultaneously should provide an experience that is most representable to actual SUD treatment exercises, as an exercise solely focused on navigation would not be of any benefit for the SUD treatment. This in contrast to separate exercises that each focus on either navigational or utilization aspect of exploration.

While the latter allows us to better distinguish the needs and implications of use in SUD treatment for either aspect separately, we believe that providing the more representable experience in which both aspects appear simultaneously allows us to more easily determine potential large bottlenecks in the interaction that hinder joint exploration as a whole. We think distinguishing the needs of the two aspects of exploration separately is better suited for the follow-up iteration to determine how to refine the interaction even further. As such, we chose for an exercise in which a treatment provider and client have to navigate through a city which contains various SUD-related encounters that need to be handled before or while continuing to navigate.

To better understand a participant's experience throughout the exercise, they are asked to make their intentions with the solution verbally clear in a manner of think-aloud, for example "I want the client to go left, so I point to the left with the EPU avatar". A researcher is present in the room to take note of these comments and observe any other notable actions performed by the participants which are not verbally declared as well.

Furthermore, to gain additional insight on the participant's experience, the exercise is followed by a semi-structured interview. The questions of this interview are designed to incite relevant experiences and reflect on these experiences in the context of the different evaluation questions. Making the interview semi-structured allows us to deviate to topics that we did not consider relevant during the interview.

Before the start of the exercise, participants should be instructed on the interaction possibilities the prototype provides. However, throughout the exercise they should not be able to ask questions on how to approach certain situation given the interaction possibilities to avoid bias in use.

4.3.3 Equipment and layout

The experiment was ran using a desktop PC[11] including two mice, a monitor, and dedicated microphone as well as a first generation HTC Vive (2016) setup. The experiment took place in a room of roughly 16m^2 at Tactus Verslavingszorg. A full overview of the equipment used, layout, and picture of the setup can be found in Figure 25. Note that the second extra mouse for the researcher was included for hygienic reasons only and was only used for setting up the experiment and setting up the interview recording during the experiment. Lastly, a summary of all digital third party resources used for the realization of the prototype, such as Unity packages, images, or sounds, are summarized in Appendix D.

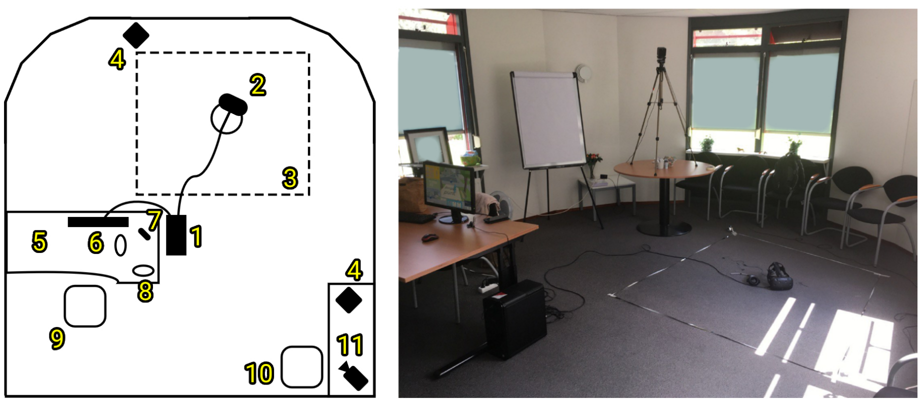


Figure 25: On the left a graphical overview of the layout all equipment used for the first iteration evaluation: desktop PC (1), IVR HMD (2), outline of VR play area (3), VR beacons (4), desk (5), monitor and mouse (6), dedicated microphone (7), extra mouse for researcher (8), chair for participant (9), chair for researcher (10), and video camera (11). On the right, a picture of the physical implementation of the setup.

4.3.4 Procedure

Participants were invited into the experiment room and introduced to the researcher and research assistant. After physically signing the consent form, the experiment could start. If participants had no prior experience with the concept of IVR, a brief verbal introduction was provided by the researcher. This instruction was supported by a short video of the research assistant looking and teleporting around in a virtual city using the IVR equipment.

Thereafter, the participants received a verbal introduction of the room setup, reiteration of the flow of the experiment, and an explanation of what is expected from them as participants during the exercise, which is that they are encouraged

[11]Desktop PC specifications: Windows 10 Pro, Intel(R) Core(TM) i7-7700K CPU @ 4.20GHz, 16,0GB RAM, and NVidia GeForce GTX 1080 graphics.

to explain their intentions with the system aloud and that they cannot ask the researcher for help with regard to the contents of the exercise.

Then, the components of the prototype and their purpose are explained: the mouse is used to interact with the dashboard shown on the monitor and that the dashboard shows all possible functionalities. Thereafter, participants get time to read through the build-in manual and try the functionalities for themselves. After a few minutes, the researcher handles any remaining questions and show any functionality that the participants did not try out themselves. Upon finishing the explanation, the actual exercise together with the research assistant could start.

Client persona

During the exercise, the IVRU role was filled by a research assistant that behave like a client according to a persona. In short, the persona is a 25 year old woman who has borderline intellectual functioning and got alcohol use disorder due to a lack of social control living alone. Her high and medium risk situation include going out with friends, boredom, and coffee in which she normally puts liqueur. She is familiar with non-IVR exercises for SUD treatment, such as role-play. The research assistant received prior instructions on how to behave at specific alcohol-related encounters during the exercise. For example, to stop and get distracted by a cocktail bar sign. During these encounters it was up to the participant on how to handle the situation. The description of the persona as it was presented to the participants and the persona instructions for each of the encounters can be found in Appendix C.

Joint exploration exercise

For the exercise itself, the general goal is for the participant to provide navigational and therapeutic guidance in the VE for a client that needs to navigate from a set starting point in a virtual city to a set destination, a bus stop. Before the start of the exercise, the participant was made aware of the location of the bus stop to be able to provide navigational guidance in a for them relatively unknown environment. During navigation, the IVRU comes across various alcohol-related encounters that are relevant to the client persona. The participant was free in how they provided guidance, given that they used the tools provided by the prototype solution.

The research assistant was instructed to adjust their action speed to match that of the participant in order to give participants time to figure out how they would want to use the prototype solution. For participants that were exceptionally quick in reaching the final destination, an additional encounter was implemented for which the research assistant was instructed to increase the difficulty of the exercise by speeding up their actions.

Figure 26 provides a full overview of the virtual city, including the alcohol-related encounters, start and end point of the exercise, and most likely routes between the start and end point. We ensured that each encounter occurs on any likely route at least once to ensure experiences between participants are similar.

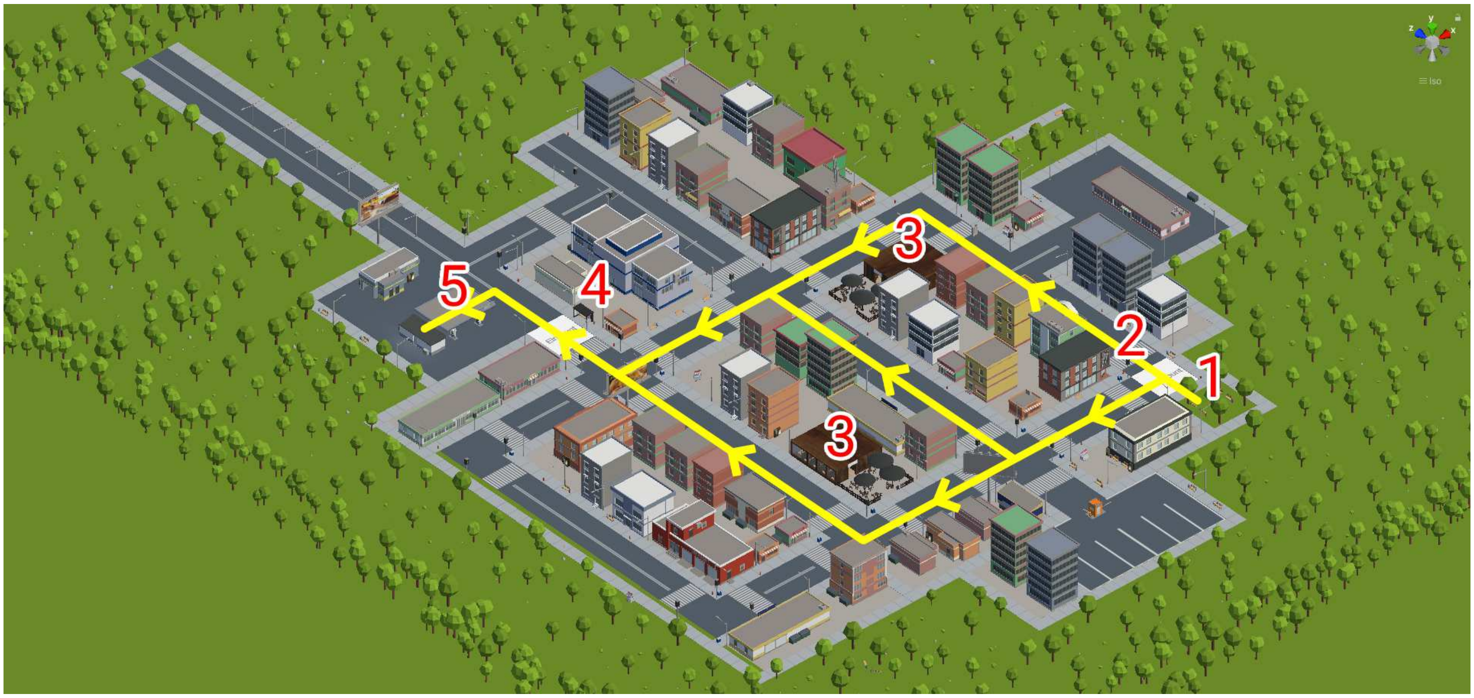


Figure 26: A visual overview of the virtual city in which the exercise is played out. The highlighted points are: park bench (1, starting point of the exercise), closed cocktail bar (2), accessible lunchroom with bar (3, 2x), bus stop (4, end point of the exercise), and gas station (5, optional encounter for quick participants). The yellow arrows indicate possible routes that participants are likely to take. Note that each route goes past every encounter at least once.

What follows is a brief description of the encounters, including their colour-coded risk-level for the client persona, with Figure 27 providing accompanying visuals.

Starting point - park (green) At the starting point, there are no direct temptations, but also no indications for where the bus stop might be. The starting point is practically on the opposite of the city respectively to the end point, ensuring the need and possibility for more complex directional guidance.

Closed cocktail bar (orange) Near the starting point, there is a closed cocktail bar with a bright neon sign as first encounter. Since the bar is closed, it is not a very risky situation regarding the client persona, making it a relatively easy first encounter to get more of a feel for how to use the application. However, there are not many distractions or alternatives available in the direct vicinity, making the participant rely mostly on themselves for handling the encounter.

Accessible lunchroom with bar (red) Further into the city, one of two accessible lunchrooms with bar is encountered. This is the most prominent encounter, as it is a high risk scenario for the client persona. The building contains various alcohol-related temptations, such as beer bottles and wine glasses. In addition, there are alternatives and distractions present, such as soft drinks or a smartphone as well. Additionally, the close quarters in the building challenges the participant to keep the camera in a clear overview position of the situation to get the most out of it. The combination of having to handle the high risk situation in a small space is the peak challenge of the exercise.

End point - bus stop (orange) After the peak encounter, it is only a short distance further to reach the bus stop, however, the exercise is not finished on arrival. At the bus stop, the bus has yet to arrive, making the client persona wait. Waiting is a medium risk situation for the client persona. The participant suggesting to take distance is not a viable approach here, because that might cause the client to miss the bus, forcing the participant handle the situation differently.

(Optional) service station (red) This optional encounter is used if the end point was reached within 15 minutes. The service station is an encounter similar to the lunchroom with bar as in there is a temptation, coffee, and various alternatives in a tight space. For this encounter, the research assistant is instructed to increase their action speed to make this encounter even more challenging for those who managed to pass the previous encounters with ease.

City in general (green/orange) Throughout the city are various minor temptations, distractions, and alternatives present as well that the participant can use in their guidance. For example, advertisement billboards, snack and drink machines, and ice cream stands. Visual examples of these elements can be found in Appendix C as well.

During the exercise, the researcher took note of the time and the exercise was stopped upon reaching one of the following conditions: the maximum time of twenty minutes was reached, the end point was reached after more than fifteen minutes, the optional encounter was finished, or the participant reached at least the lunchroom with bar and they tried their best to utilize the available tools, but progression of the scenario stagnated due to the discrepancy between intentions of the participant and ways provided by the prototype to answers these intentions. Throughout the exercise, the researcher observed the participant's behaviour, documenting noteworthy interactions that could be cited during the semi-structured interview that followed. After the exercise was finished, the research assistant left the room such that the participant could speak openly about the assistant in the interview.

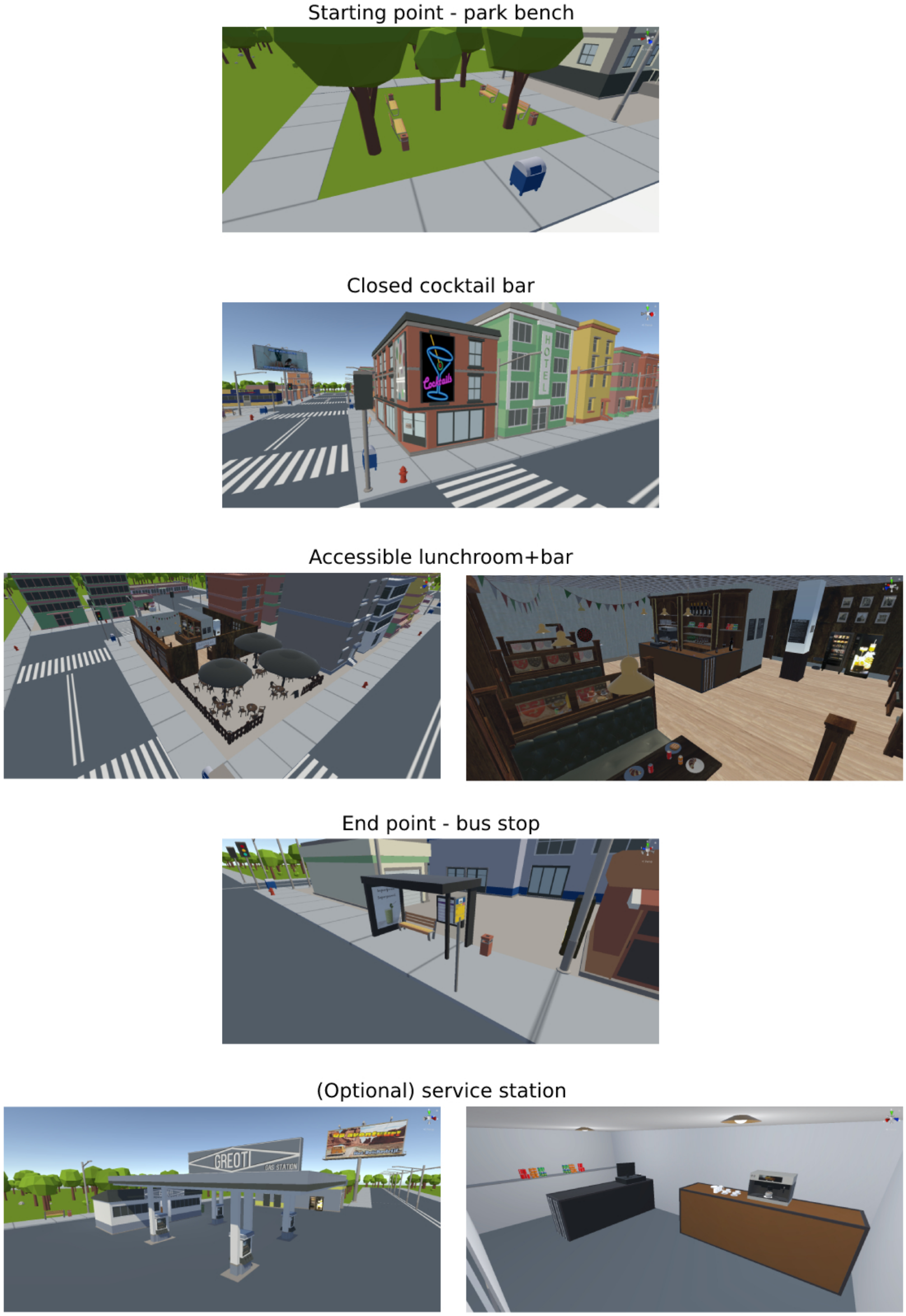


Figure 27: Encounters as presented in the VE during the exercise. From top to bottom, the encounters are shown in order of how they are encountered during the exercise.

Semi-structured interview
After the exercise followed a semi-structured interview about how the participant experienced use of the prototype during the exercise in relation to the evaluation questions. Table 5 provides an overview of the questions that were used during the evaluation and how they relate to the evaluation questions. Note that the original questions are in Dutch and can be found in Appendix C.

4.3.5 Processing and analysis

The interview results are processed using inductive content analysis to ensure that any topics that were not identified to be relevant beforehand can be easily taken into account as well. The inductive process consists of three main steps: fragmenting (*i.e.* coding), grouping, and finding structure. Due to the interview being semi-structured, part of the coding categories can be determined *a priori*, whilst the rest should emerge during the coding process.

4.4 Results

4.4.1 Participants

The experiment was performed with five different participants consisting of two male and three female participants between 24 and 42 years old (Mean: 32, SD: 7,267). The participants will be referred to as participant 1 through participant 5 or P1 through P5 for short. Of the participants, only participant 1 and 4 had prior, although limited IVR experience. Participant 5 explicitly indicated that he is not very technologically savvy himself. All participants finished the exercise by getting the IVRU to the designated location in the VE and all participants were quick enough to include the optional destination, the service station, in the exercise.

4.4.2 Results

As certain results are covered by multiple codes, the results are grouped into three main topics based on the evaluation questions, as opposed to describing the results per code. The three main topics are: *joint exploration*, *presence and social presence*, and *use in therapy*. For readability, the results are further grouped into related sub-topics. The full coding structure is provided in Appendix E. Full transcriptions, including codes, are available to other researchers upon request by contacting the corresponding supervisors of this research.

At various moments during the interviews, participants indicated that certain struggles with the program were caused by inexperience, rather than calling them flaws outright, and that their view might differ once getting more used to the application.

Joint exploration
Individual viewpoint EPU

Table 5: Translated interview questions of the semi-structured interview for the first iteration evaluation

Related EQ	Interview question
EQ1	What is your first impression of the system?
	How did you think the exercise went yourself?
	Was the system easy to use?
	Were there certain options within the system that surprised you?
	Were there certain options that were better than the others?
	Was there something that you were clearly missing in the application?
	Do you think the system was generally designed for use by treatment providers?
	What aspects do you think need to change before you would use the system in practice?
	Do you foresee problems with the system when using it with real clients?
	How would you use the system differently when using it with real clients?
EQ2 & EQ3	How do you think the communication between you and the VR user went?
	Could you easily understand the intentions of the VR user?
	Did you have the feeling that by using the system, your intentions reached the VR user?
	(If prior experience with asymmetric VR) How was the communication different than that earlier experience?
EQ4	Did you have the feeling that you were also partially present in the virtual world?
	Or did you have the feeling that the virtual world was partially present in the real world?
	Did the VR user and their representation, the blue figure, feel like the same person?
	Did you feel connected to the orange figure?
	Did you have the feeling that you were both together in the virtual world?

Participant 1, 2, and 3 explicitly appreciated that they had their own individual viewpoint in the virtual environment. Participant 3 and 5 recognized the overview perspective in the main window and used it to better anticipate on potential risk scenario's for the client and determine where to go next respectively. Participant 5 added that in this perspective it was sometimes challenging to imagine how certain aspects of the environment would look like from the IVR perspective, especially estimating virtual distance between users: "You also really don't know the distance. [...] Because how far are we actually standing away from each other?".

Participant 2, 3, and 4 indicated that they would omit their individual perspective and fully focus on the client's point of view whenever there is less need for guidance: "Because you know, when watching through the client's eyes, you know where you roughly are in the world and that you can find the bus stop from whichever point you are" (P3) and "because you're not there in the world anymore, then you don't need that anymore. Then you are fully focused on what the client is seeing and doing" (P4).

Getting a feel for the camera controls seemed to be an aspect of the application all participants struggled with most, but was often only briefly touched upon during the interviews. For example, "I noticed that with [name client persona] I was sometimes messing around with the screen" (P1), "But because it is the first time, you are really looking for 'when do I need to press which button'" (P2), and "[...] I was busy with myself a lot. Yes, you know, I was clumsy, [...]" (P5). The participants expected to get better at it with more experience: "I think it is purely habituation" (P2) and "[...], but that is just practice" (P4).

The *follow position and manual rotation* camera option remained unused in most experiments. Participants simply used the *fully follow* option when they needed the client in view and occasionally switched to manual controls to look around themselves or perform an action. Participant 1 noted that the *follow position and manual rotation* option might have been useful in some cases, but she indicated to have forgotten about it during the exercise.

Furthermore, participant 3 recognized the different zoom levels and zooming out to a map-like perspective felt logical. He suggested to add a compass to support determining the orientation of the camera as well. Participant 4 noted that using arrows keys on a keyboard, as opposed to using the mouse, to move the camera position would have been easier and more logical for her.

Awareness cues

Participant 1 indicated that she found it useful that she could check how certain visuals looked from the client's perspective using the WYSIWIS window. Participant 2 had a similar opinion and adds that she used the WYSIWIS window to determine the orientation of the client respective to her overview position "I didn't really pay attention to the blue figure, because I was only looking at that screen into the world for orientation, to what she was seeing and how she is standing there". Participant 4 had an opposite view, as she was in favour of not having the WYSIWIS window visible, arguing that the IVRU avatar provided similar information well enough "[...], because now you see in your screen the figure of the client as well, so you can imagine roughly what they are seeing... for me the screen was superfluous". Notably, the viewing direction beam and field of view frustum cues remained unused among all participants.

Verbal communication

All participants indicated that verbal communication between them and the IVRU was sufficiently clear, where participant 1 explicitly indicated it was due to being in the same room as the IVRU. For participant 5 it was clear most of the time, but sometimes the IVRU spoke softly, making her harder to understand. Participant 2 and 4 mentioned that verbal communication is key in such exercises, as conversations about what the client is experiencing are prominent during the rehabilitation process "you simply do a lot in conversation form" (P2). This setup enabled such conversations well enough.

Participant 2 indicated that she thinks that verbal communication via the EPU avatar is not necessarily needed. She thinks that a disembodied voice might be sufficient, as long as it is properly explained beforehand to the client. She adds that she did not find it weird that clients might be turned away from her during conversations, as she knows that the client is having a conversation with the EPU avatar "Because you know that they are actually talking to the virtual figure it's no problem that you are talking to someone that seems not to pay attention".

Communicating intentions by the EPU

All participants except for participant 3 indicated that their intentions would come over sufficiently well using mostly verbal communication and occasionally supporting it using the provided actions. Participant 3 doubts whether the intention of emotional reflect and validation comes across well. For example "impressive of you" accompanied by the applause action or "still difficult for you" for support and sympathizing respectively.

Furthermore on emotional reflection, participant 4 indicated that she conveyed praise only verbally and that she could have applauded when the IVRU left the bar, as "it is very impressive by the client, especially them having a lower IQ". More specifically for the applause action, participant 1 simply thought it was a nice detail, whilst participant 3 and 5 indicated that they would not use the applause button at all. Participant 3 found using the applause button to convey his praise would be too artificial and participant 5 would only applause on exceptional occasions. In the same light, participant 5 indicated that additional gestures, such as a beckoning gesture or a thumbs up, would be helpful in specific scenarios where the current gestures would fall short. Participant 2 indicated to believe that live contact is more useful than trying to get the point across in the virtual environment "For me, live contact is more important than contact in that world", and that simply using his voice, without supporting it by actions, would suffice most of the time.

Participant 5 noted that clients might get scared because they are essentially blindfolded. Participant 5 recognized that the EPU avatar being there, using the actual voice of the EPU, might help ease the mood despite its simplistic appearance. Participant 2 indicated that she thinks she could ease the mood of clients because of the well enabled verbal communication. Both participant 2 and 5 mentioned that humor is a nice tool during communication and that they could convey humor well enough with the current solution.

Furthermore, participant 2 indicated that she would probably not be as prominently present with the EPU avatar and would therefore mainly use the general environment actions to visually support communication. Participant 4 indicated that for this particular exercise the temporary landmark supported by the attention grabber would suffice, but for other exercises, a combination of other options might be required. She argued that the temporary landmark was nice for specifying destinations in the larger city environment, but not so much for smaller areas, such as the bar, where the outline or the attention grabber would be better suited. Participant 1 was unsure about the use of the attention grabber when pointing with the EPU avatar would suffice. For using actions in general, participant 4 suggested that certain actions should be turned off after using them once, such as the temporary landmark, in order to not accidentally use them again.

Notably, participant 3 and 5 initially indicated that they would have preferred to use IVR technology as well. Participant 3 would want to use IVR technology to help apply his body better for conversation techniques and participant 5 stated that he could then more easily perform actions in the environment. However, both participant 3 and 5 stated that they were unsure whether they would want to give up their non-immersed perspective and overview position in the current application in exchange for these mentioned needs.

Understanding the intentions and feelings of the IVRU

Participant 2 and 4 indicated that the intentions of the IVRU were clear, as it was clear on the screen what the IVRU was going to do and the IVRU answered questions about her feelings or why she did something. Participant 3 noted that to him the intentions were not as clear regarding what the IVRU was going to do when encountering risk scenarios “I didn’t know what her plan was”. For example, participant 3 indicated that the IVRU lacked an observable doubting gaze towards the bar before she was going to enter it. He adds that the IVRU’s intention might not have been to go drinking in the bar, but he lacked the possibility to observe a “I would like to drink that” reaction for example. He recognized that for example a button that the IVRU can press or facial recognition to indicate cravings to the treatment provider on the dashboard might help. Participant 5 mentioned that he did not expect the IVRU to enter the bar, hence he did not anticipate on observing any indications of intentions that would lead to entering the bar.

Furthermore, all participants noted the importance of being able to observe body language of the client when they are experiencing triggers, so that they can respond to these reactions. Participant 1, 3, and 5 explicitly indicated that they would have looked more towards the real client if they did not have to focus as much on the desktop monitor as they did currently. Having to focus on the monitor was indicated to be caused by inexperience. Participant

5 added that clients often communicate more with their body language about how they feel, in contrast to the research assistant IVRU who communicated it verbally. Participant 2 indicated that it could be possible to only observe the physical behaviour of the virtual representation, but only if it perfectly mimics the physical behaviour of the client.

Similar to recognizing cravings, recognition of emotions and displaying those on the dashboard was recognized to be a useful addition by participant 3, however, he indicated that the additional necessary technology might make the setup too complex "then you gear up a lot for a dry run". Participant 5 is also in favour of an indicator on the dashboard, such that he would notice it when being too focused on the monitor.

Creating a common point of reference and redirecting attention

Participant 1 and 4 explicitly recognized the ability of creating a common point of reference using the verbal communication supported by the tools within the program. For example, placing the temporary landmark and asking "do you see the flag?" or pointing to a building asking "do you see the orange building?". Participant 4 added that this redirection of attention could also be effective when the client is distracted by elements of the virtual environment that are not part of the exercise. Participant 3 on the other hand indicated that during the exercise he did not seem to have the tools available to redirect the attention of the client towards the EPU avatar. In hindsight, he indicated that the attention grabber might have helped.

Participant 5 thought the wave action was a logical option to go for when needing to redirect the focus of the client towards the EPU avatar "'Hey, do you see me? Hello!', very friendly". In contrast, participant 2 indicated that she did not know what she would ever use the wave gesture for. Participant 5 indicated that a different redirection approach is probably needed for every client, which also depends on the familiarity with a specific client, and he would mostly use verbal cues in that regard. On the other hand, participant 1 indicated that she requires the ability to point to objects and walk along with client as well "That you can talk, point, and walk along. Those three" and that those three were much more essential than the general environment actions.

Naturalness of performing actions and interaction

Participant 1 and 5 indicated that they would have sometimes preferred to perform actions using IVR technology, such that they would feel more natural, as opposed to pressing buttons on the dashboard. Both indicate that mimicking real life interaction as much as possible is a good thing to strive for when further developing the system. Yet, participant 1 indicated that she was unsure whether she could better use the system if virtually present due to the lack of an overview position "There are pros and cons: if I would have been there, it might have felt more natural, but this is better for getting an overview". Participant 2 indicated that performing actions using buttons was less natural than performing them yourself, but the resulting actions mimicked real life interaction sufficiently. Similarly, participant 3 indicated that using buttons to perform actions feels unnatural, but that should not be a reason not to press them for practical actions. For example to give praise, he would rather give a pad on the shoulder, a wink, or fist bump for real, as converting these gestures to action buttons would still feel “artificial”.

Participant 4 indicated that with more natural interactions, she might feel a stronger connection with the EPU avatar if it mimicked those natural actions “Yes, perhaps it becomes more ‘your’ character so to say”. Additionally, Participant 1 indicated that the unnaturalness might be due to the fact that actions have to be performed sequentially. For example, select a destination to walk towards and then point towards a specific object, as opposed to starting to walk and point simultaneously. Regarding the input methods used to perform actions, participant 5 suggested that using touch input might feel more familiar and natural compared to mouse movements as touch screens are more and more prominent in today’s society.

Participant 3 and 4 indicated a certain experienced distance when communicating via the system compared to face-to-face conversations. Participant 4 notes that having this distance in communication, having a conversation that way might simply not work with every client. Participant 3 compares the experienced distance to talking on the phone “Its more comparable with calling someone than that you are walking together despite being in the same room”.

Knowing each other’s whereabouts

All participants except for participant 3 indicated that they sometimes lost track of the whereabouts of the IVRU avatar when they temporarily focused elsewhere “Sometimes I stayed behind and they were already gone” (P2). More specifically, participant 5 indicated that when he took the lead and the client strayed off it was hard to readjust “I turn around and she is gone. I see where she is. But at a certain moment I just failed... I think it is a matter of habituation... then I have to turn around and wave again”. In similar fashion, participant 1, 3, and 5 indicated that they sometimes lost track of their own EPU avatar as well “At a certain moment I walked behind the bus stop and I thought ‘where am I’, looking up, ah yes, there I am and you’re back. Purely habituation” (P5). Participant 3 suggested visualizing the avatars whenever they are out of view or behind an object, using an outline or arrow for example.

Usage patterns navigation

To complete the navigation part of the exercise, patterns in application-use varied widely between participants. The most common approach was used by just two participants, 3 and 5, which was letting the EPU avatar take the lead manually, as they “knew the route” and walk back whenever the client stopped. Participant 3 indicated that in real life, whenever the client knows the route, he would follow them instead. Participant 5 added that he would rather have the avatars walk next to each other, similar to real life, as opposed to one following behind the other.

The other participants individually used the following patterns. Participant 1 let the EPU avatar mainly walk along the client to be able to focus on where to go next, using solely EPU avatar actions to support verbal communication. She did add that when she had a reason to use the manual walk action, she had forgotten about it. Participant 4 mainly used the temporary landmark to create a visible route for the IVRU to follow. Lastly, participant 2 mainly gave verbal instructions based on the client's point of view without using any actions for visual support.

Regarding their approach, participant 3, 4, and 5 indicated that the exercise being specifically 'get the client to the bus stop' influenced their way of guidance. Participant 3 and 5 indicated that having such a specific task "made it some sort game" for them as opposed to a treatment exercise for the client.

User interface appearance
Participant 3 and 5 noted that the user interface, despite its many options, was clear and did not diminish the overall experience. Respectively, participant 3 and 5 mentioned that the plain look made it coherent and clear and that the chosen colours and rounded edges made it look user-friendly. Participant 2 mentioned that the user interface sat not in the way of observing the virtual environment. Lastly, participant 4 indicated that it was sometimes hard to focus on two screens, the main window and IVRU WYSIWIS window, at once. She added that she would have turned off the IVRU WYSIWIS window if she had remembered that she could, as the awareness cues in the main window should suffice in her opinion.

Presence and social presence
Spatial presence EPU
All participants except for participant 5 indicated to not feel present in the VE "At some point, I did not even realize she was standing there. So I really was inside of the world" (P5). Neither did the VE become a part of the real world for participant 1, 2, and 3. Participant 2, 3, and 4 specifically indicated that they found this lack of presence not to be an issue. For example because they preferred the overview position as mentioned by participant 3. However, participant 3 did indicate that occasionally they would feel more present when having to focus on the VE during risk situations "During risk situations I did, especially when we were in the coffee bar... Then you have a realisation of 'okay, I need to focus, because the client is in a risk situation'".

Virtual representation EPU via avatar
All participants indicated that they did not have a particularly close connection to the EPU avatar. All participants except for participant 2 indicated that they felt more like the pilot of the EPU avatar rather than being the avatar or having a more intimate connection with the avatar. Participant 2 and 4 specifically added that the lack of connection is not a problem, as long as the IVRU experiences the EPU avatar as a representation of them. Participant 2 noted that they did not connect with the avatar and that it did not add value to the experience because it was too "weird". Participant 5 mentioned that this

lack of connection might be due to inexperience with the application. However, participant 1 indicated that having more application experience would likely not change her view. Participant 4 suggested that if the EPU avatar would have displayed more mimicry, as the IVRU avatar did, it might increase the connection "Then it becomes more 'your' figure, but she is unsure whether a more intimate connection is possible. Similarly, participant 3 was unsure whether adding more mimicry from the current viewpoint would increase this connection, but he indicated that having a first person's perspective in addition to more mimicry would definitely help.

For participant 1, 3, and 4 this lack of connection would not change if the avatar would look more like them. All participants except for participant 4 did recognize that a change in appearance might add value for the client. More specifically, participant 2 mentioned that the EPU avatar would only have value with a more realistic appearance and no value otherwise. She adds that having the avatar look like herself would probably be distracting for her and instead suggested to use a "common male or female figure", however, this would not influence their feeling of being present in the VE. On the other hand, participant 5 indicated that having the EPU avatar look like him would increase his feeling of being present in the VE. He adds that a more realistic appearance would add to the experience of the client as well.

Furthermore, participant 1 noted that the EPU avatar's current appearance, added voice, and actions should be enough representation for clients to understand who it should represent "So that they really understand that it is I who is standing next to them". Participant 1 thinks that incorporating a webcam in the setup, for example that the head of the avatar becomes a screen displaying the webcam image, would probably make things more disturbing for the client. She argues in favour of the current appearance "precisely because it clearly represents a person and nothing more".

Moreover, participant 2 mentioned that certain unnatural movements that were performed by the avatar were a considerable distraction for both them and likely the IVRU and as such would make it harder to keep the exercise serious. For example when trying to let the avatar point and applause simultaneously "that figure was doing some really weird things" (P2). Participant 3 was uncertain about how clients would receive the orange appearance. However, he indicated that for them as treatment providers, the abstract representation should not be a problem.

Virtual presence IVRU as perceived by the EPU

Participant 2 and 4 indicated that they felt that the IVRU was present in the VE. "I know that she sees an entirely different image and that she is inside there" (P2) and "you notice the experience of the virtual environment. [...] She is really inside of that world" (P4).

Virtual representation IVRU via their avatar

All participants except for participant 2 indicated that the IVRU and IVRU avatar felt like the same person, but all for different reasons. For participant 1 it was related to who she focused her cues towards "looking at the screen, you ask the blue figure to do things, rather you ask the client, however, the blue figure acts on it". For participant 3 it was not because of any mimicry, they just accepted it "I just know. You says it represents the client and I think 'sure'". Participant 5 indicated that this was because he could interact with the blue figure as if it was a real person "I really saw her, when looking at the figure". Participant 2 indicated that she did not really pay much attention to the IVRU avatar aside from determining the orientation of the client within the VE.

Furthermore on the appearance of the IVRU avatar, participant 1, 2, and 5 indicated that for them as treatment providers, it does not matter that the avatar is a simple blue figure. Participant 1 and 2 explicitly added that making the IVRU avatar more realistic would not add to the experience.

Social presence and taking on the VE together

Participant 1 thinks that with more experience she would have had more the feeling that she was using the system together with the IVRU, as opposed to being a distant pilot. However, she added that this is already an improvement on interaction between users in a conventional IVR setup.

Specifically for guidance, participant 3 would have preferred to feel social presence. He currently did not because of the discrepancy between their positions in real life and the VE "it's more the fact that I am sitting here and she is standing over there, while you are walking together in the environment. In my mind that is not right". He added that he had the feeling that he were for 25% in the virtual world and for 75% in the real world and that he felt that the client was for 75% in the virtual world and for 25% in the real world. He argued that this is the case because he knew that he was using the same program as the IVRU, but he kept being aware of being behind a desktop monitor most of the time.

Participant 3 and 5 indicated that they did have the feeling that they were taking on the virtual environment together with the client. Participant 5 indicated that this came to be because he could communicate with the IVRU about the VE. In contrast, participant 4 indicated that she did not had the feeling of taking on the world together and that she was "watching along at a distance".

Use in therapy

Accommodating users

Participant 1, 2, and 5 indicated that the system essentially contains everything one might need for being a treatment provider for similar exercises. Participant 2 and 4 indicated that different clients are likely to prefer different guidance tools, but they think that with the options the application provides there should be a combination that accommodates any client. Additionally, participant 2 and 3 mentioned that the application seemed to be designed for guiding people in IVR and that for participant 2 the application did not restrict her in her regular pattern of guiding clients. However, participant 2 added that the application did not seem to be designed for specifically accommodating treatment providers during that guidance as well "What you are doing during therapy is mainly asking 'okay, how does that make you feel' [...] It is designed to guide people in that world and it doesn't matter if I am a therapist who is doing it".

Simultaneously being a treatment provider
All participants found it challenging to simultaneously control the application and be a treatment provider during the exercise. For example, participant 1, 4, and 5 indicated that they noticed themselves lacking in therapeutic capabilities while struggling to get the application under control "So getting the technical part under control was a getting used to, but in the meantime the IVRU has cravings and I have to act on it" (P1). In contrast, participant 3 indicated that he had already forgotten about certain actions he would have liked to use during the exercise but he was too busy being a treatment provider. Participant 3 and 5 indicated they would have done a better job if they, for example, knew the environment or the client more by heart, such that they could have anticipated specific high risk situations or better determine what would be the best course of action for that specific client in such situations respectively.

Break outside of IVR or turning off stimuli and discussing new strategies
Participant 2, 3, and 4 indicated that they would have the clients take off the HMD whenever the client gets overwhelmed by the virtual environment and indicated a preference for discussing the overwhelming scenario and potential strategies for handling that scenario face-to-face. Participant 3 mentioned that he does recognize an added value in having such a conversation in the VE to directly apply to the discussed strategy. Participant 4 recognized that a button that turns off stimuli or changes the virtual environment to something more calming could work if explained properly to the client in order to avoid confusion. However, the participant added, depending on how triggered the client is, it would sometimes be wisest to simply take off the HMD "before things get demolished".

Usage behind desktop
When the topic explicitly fell on using the dashboard behind a desktop setup, participant 1 indicated that she could better guide the client because she had a better overview than she would have in IVR and participant 3 indicated that he could easier observe the body language of the client. However, individual participants also indicated that being behind a desktop monitor resulted in a less immersive experience (P1), yet could sometimes miss client body language by being too focused on the monitor (P3), and she would not be able to physically intervene as easily when necessary (P4). Regarding physically intervening, participant 4 suggested the usage of a laptop to be able to sit physically closer to the client and with less obstructions, unlike with the desk in the presented desktop setup.

Usage with real clients and application as a video game
Participant 1 indicated to take much more time before such an exercise to ex-

plain the setup and the technology, as, for example, clients might not understand the IVR controller and might already get frustrated before the start of the exercise. Both participant 1 and 5 indicated that clients might become scared because the HMD cuts them off from the real world such that they cannot see what the treatment provider is doing in real life, disrupting the exercise. They recognize that this fear might be possible to overcome with substantial preparation, but even with that preparation it will simply not work for all clients. Participant 5 did note that the EPU avatar might be able to help overcome this fear.

In similar light, participant 5 indicates that he is unsure about how a client would experience knowing that the therapist is watching along whenever they are not represented in the VE. For example, it could feel like they are being spied on. Additionally, participant 1 noted that for clients it might outright be confusing if they lose track of the whereabouts of the treatment provider in real life. Participant 2 indicated that she was uncertain how clients would react to the application as a whole.

Additionally, participant 2 noted the importance that the client should be able to be alone in the virtual environment in order to really practice leave of absence. She adds that during exercises in real life, treatment providers either join from the very start, while only separating when the client has to go into the grocery store alone, or that the client goes by themselves altogether.

Participant 2 notes that she expects real clients to be less helpless than the fake client played by the research assistant and would probably require less guidance, particularly mentioning “Our clients certainly know the difference between left and right”. However, participant 1 indicated that it is not uncommon for clients to not be able to make the distinction between left and right, as well as easily following descriptors such as *the orange building*, and the treatment provider’s guiding strategy should be adapted if necessary.

Furthermore, participant 1 mentions that most clients are not used to the luxury of IVR technology. On the other hand, participant 3 indicated that especially the younger clients would not be surprised by the IVR technology or appearance of the virtual environment, as they are more likely to be familiar with video games. However, participant 3 adds that the older generation of clients is likely to have difficulties grasping the technology. On the notion of video games, participant 1 indicated that “clients might think they are part of a video game at first instead of a serious leave of absence exercise”, but after some practice this should not pose a problem.

Since it is a virtual environment, participant 4 noted, some clients might do things they would not normally do in similar exercises, such as entering a bar, simply because “they want to know how it is made virtually in the computer system”. Related to this behaviour, participant 5 mentions that clients are often willing to cooperate with the exercises, hence there is no need to focus on measures to handle unwilling clients. In rare cases that it does occur, he notes, “you simply stop the application”. Regarding the use of physical contact during interactions, both participant 4 and 5 indicated that physical contact with clients is a rare occurrence and would only be used if they are certain that it would help the client in specific situations.

Participant 3 indicated that communication with familiar clients would probably go smoother than with the unfamiliar fake client played by the research assistant. Lastly, participant 1 and 4 explicitly mentioned that they think clients will enjoy practicing using this setup.

Virtual environment
Although not explicitly being part of the interview, participants were eager to comment on the design of the virtual environment in which the exercise took place. Participant 1, 2, and 5 noted that the environment contained the right triggers and basic high risk scenario's, such as bars, a gas station, and a park, and participant 2 praised how the bar was presented in more detail than the rest of the environment. However, participant 2, 3, and 5 mentioned that the environment needed to be more dynamic, such as other people or traffic, in order to better mimic real life situation and thus be better practice material "There was nobody on the streets... You get a bit of a Sunday morning feeling" (P3).

Additionally, participant 3 indicated that the environment as a whole was relatively small. Participant 1 indicated that more interactions with the environment, such as being able to purchase a drink, would be needed to improve the current environment. Participant 5 mentioned that stress resulting from a more alive environment would also elicit cravings in general and he adds that he would like to tailor the environment more towards the specific triggers of clients, such as a specific type of music.

Placement within the treatment process
When asked about the placement of this system within the treatment process, participant 1, 3, and 4 indicated to implement it early within the process, before clients' first actual leave of absence. They reason that it can help them determine what a clients triggers actually are (P1), it is a nice preparation on the actual high risk scenarios without having to consider all other aspects of real life simultaneously as well (P3), and it can help them determine whether clients are prepared for real leave of absence (P4) respectively. However, participant 1 notes that it is certainly not a full replacement as in its current form the scenarios are too simplified. For example, the scenarios do not cover social ambiance or smell.

4.5 Discussion

With the first iteration evaluation we aimed to answer the four established evaluation questions. We will discuss relevant results for each question individually, as well as results that we deem noteworthy to take into account for the second iteration and that are not covered by any of the evaluation questions.

4.5.1 Critical aspects for SUD treatment

The results indicate the following for EQ1, *How accommodates the solution critical aspects that an AVR solutions needs to adhere to to make that solution suitable for use in SUD treatment?*

Verbal communication was indicated by all participants to be crucial for SUD treatment and that in the current solution verbal communication was enabled sufficiently well. On the contrary, there seems to be a lack of visual cues to communicate affect, specifically for conveying emotional reflection and validation. For example, giving praise or sympathizing, as was indicated by two participants. Adding more gestures to accommodate these missing gestural cues in the current approach can be a solution, such as the suggested thumbs up or pat on the shoulder. However, it was also made clear that a specific gesture might not convey the correct message. This was for example made clear by participants that did not want to applaud to give praise. Accommodating the many different ways to convey certain intentions by adding more gestures would essentially require an indefinite number of buttons. Therefore, the underlying problem seems to be the lack of a more robust way of providing gestural cues to support conveying affect in the current prototype.

On the topic of visual cues, the physical behaviour of the IVRU was indicated to play a substantial role for SUD treatment by all participants as well. Two participants specifically indicated that how the IVRU felt about the VE was made clear well verbally. However, actual clients might not be able to express themselves that well verbally and such treatment providers have to rely on perceiving physical cues. In that regard, three participants indicated that they would have looked more towards the client in real life if they were able to focus less on the monitor. They indicated that would they have more experience with the AVR solution, they might need to focus less on the monitor. However, due to the static position of the EPU in the AVR setup, it could be that the IVRU is positioned in such a way that it remains hard for the EPU to perceive physical behaviour despite not having to focus on the monitor. Therefore, we deem it beneficial to look into this aspect in the second iteration as well.

Furthermore, participants indicated they found it hard to use the prototype and simultaneously be a treatment provider. This was indicated to be the case due to them not fully knowing their way around the application and having limited survey knowledge of the environment. This in contrast to the expected use, where the treatment provider has full knowledge of the application and the virtual environment. However, this result is not entirely unexpected given that the participants were only given a brief moment of time before the start of the exercise to get themselves familiar with the functionalities and environment.

Lastly, two aspects that were made clear to be not as useful to consider for use in SUD treatment were that of accommodating clients that are too overwhelmed in the VE, for example via a virtual break room, or incorporation of features to deal with recalcitrant clients, such as blocking or redirecting movement of the client. In both cases the exercise is likely to simply stop.

4.5.2 Interaction for joint exploration

Continuing to EQ2 and EQ3, *How are interactions regarding joint navigation experienced by the users?* and *How are interactions regarding collaboration experienced by the users? respectively.*

In general, most participants think that the interaction possibilities the prototype provides are sufficient for similar exercises and that the variety in possibilities makes it so that varying individual needs or preferences of clients can be accommodated with the solution. However, we noticed that during the exercises the variety of interaction possibilities were mainly used for the navigational aspect of exploration and that SUD treatment related interactions were mainly done verbally, for which we already identified a lack of support for gestural cues. For example, creating a common point of reference and redirecting attention is relevant for both aspects, but mainly used for the navigation aspect.

The variety of available approaches was positively supported by participants in the sense that including a flexible way of providing cues can be beneficial given individual needs and preferences of users. This in contrast to determining one optimal approach for time-based tasks.

What we noticed across participants as well is that after placing their avatar for interaction, the avatar was not recalled to let the client explore by themselves at all. This could be due to how the exercise was set up, including the specific instructions of the research assistant that kept participants engaged, or the way treatment providers interact with clients that are relatively unknown for them and thus not know their level of independence for example.

On performing interactions in general, the mouse based approach seems to be a valid approach that was not specifically disliked, although one that was indicated to require some getting used to. Additionally, participants did comment on how this way of interacting felt relatively unnatural or artificial given the context of SUD treatment. Reasons varied from having to sequentially input all actions compared to doing them simultaneously in real life, to having a disconnection between an avatar that interacts for you and the camera view, to the lack of being able to provide emotional intentions via the available actions. While not indicated to be a necessity for interaction, we think looking into the artificiality can be beneficial for the second iteration as it was raised as a concern. For example if the interaction changes fundamentally, we can explicitly refer to the naturalness of interaction during the evaluation.

Moreover on the individual view of the EPU, participants indicated that having the overview perspective of the VE was appreciated during the interaction. However, when there was no need for interaction most participants preferred to watch along using the WYSIWIS window, which was deemed to provide enough information compared to the additional awareness cues provided in the main interaction window. The additional gaze direction beam and field of view frustum appear to be superfluous for joint exploration in this context, as they were never used and not missed by participants, despite being shown as options before the exercise.

In light of awareness cues, having a detachable view and separately controlled avatar, the participants sometimes lost track of the IVRU avatar's location or their own avatar's location. The application includes ways to deal with such situations, for example by switching camera modes or recalling and placing the EPU avatar again. Participants not using these ways indicate a lack of familiarity with the application and more training would help in that regard. However, it would be better to prevent those situations from happening in the first place by, for example, including additional positional awareness cues, such as highlighting obstructed avatars or have a pointer to avatars that are outside the EPU's view.

4.5.3 EPU presence, social presence, and avatar representations

For the last question, EQ4, *How were the notions of presence and social presence experienced by the users?* While presence was determined to be a crucial aspect for the IVRU, the lack of spatial presence participants experienced using the desktop was indicated not to be an issue for the interaction. One participant did indicate to feel spatially present and did also indicate that this had implications on how well they could keep an eye on the IVRU in real life. Losing the IVRU out of sight, for example by being immersed in the application, potentially conflicts with the earlier identified need to observe any physical behaviour of the client that is not represented in the VE. Being spatially present, however, did not seem to substantially change the behaviour of that participant compared to the other participants who indicated not to be spatially present.

On social presence, *i.e.* the experience of being with another during a mediated social interaction, participants were not as vocal, but the general consensus seems to be that social presence was limited. This is likely due to the interaction of the IVRU being directed towards the EPU avatar, which is observed as a separate entity due to the exocentric perspective of the EPU. As such, the exocentric perspective might have hindered Psi from occurring for the EPU, *i.e.* they might have felt left out of the interaction because the interaction did not directly relate to them, but to the EPU avatar. Yet, the interaction remains a joint experience, but with a perceived distance between users. "Its more comparable with calling someone than that you are walking together despite being in the same room", as one of the participants mentioned.

The virtual representation of the EPU avatar plays a substantial role for social presence as well. Participants indicated to feel agency over the avatar but lacked ownership in the sense that they felt they were the avatar's pilot but did not further associate themselves with the figure. However, this lack of personal connection with the avatar was not deemed a problem for the avatar's current purpose of providing a point of interaction for the IVRU. Yet, participants did indicate that with actual clients, the appearance probably would have to change to appear more human. Changing or customizing the avatar appearance comes with the additional benefit that it might help EPUs associate themselves more with the avatar as well.

While not indicated to be necessary for effective interaction, the lack of social presence does have considerable implications on the joint experience and should therefore be taken into account in the second iteration.

4.5.4 Limitations

As indicated by one of the participants, the interaction possibilities did not seem to be specifically designed for use by treatment providers. The latter comment is not unreasonable given the modest involvement of treatment providers in the requirements engineering phase. However, this idea could also come forth from how the evaluation exercise was set up, as we did not specifically base our exercise on existing treatment protocols. The main goal of the exercise was simply to get the IVRU to their destination with the SUD-related encounters essentially being hurdles. This seems in line with participants indicating that getting the client towards the bus stop felt like some kind of game. Having dedicated exercises for both the joint navigation as well as collaboration aspects of exploration could provide more insight on the matter and therefore should be considered for the evaluation of the second iteration.

Furthermore, while participants were able to provide substantial insight on actual clients taking on the IVRU role, there were uncertainties as well. Uncertainties for example about how certain interactions would be experienced as the IVRU. This indicates a need for participants in the IVRU role as well, and if not clients then at the minimum people uninvolved in the research.

Concluding, in this chapter we provided a description of the first iteration prototype, prototype evaluation, its results, and the discussion of the results. We gained more insight on RQ1 and the accompanying SQs by answering the accompanying evaluation questions and were provided direction for the second UCD iteration.

This direction includes addressing the way the EPU provides gestural cues, perceives the physical behaviour of the IVRU, experiences social presence, and becomes more spatially aware of the avatars. Additionally, separation of the navigation and utilization aspect of exploration, and experience of the IVRU, can be considered for the second iteration evaluation as well.

In the next chapter we discuss this second UCD iteration as a whole, starting with the implications of these results on the potential improvements on the prototype.

5 Second iteration

Having obtained possible design directions from the first iteration evaluation, we can start with the second UCD iteration, which is the last iteration in this research. The aim for this iteration is to improve the first iteration prototype using the results of the first iteration evaluation to help us get closer to answering the main research question. Similar to the first iteration, we start with establishing requirements and brainstorm how to accommodate these requirements, followed by a description of the second iteration prototype, with a description of the evaluation and its results thereafter. The discussion of these results relative to the research as a whole together with the limitations of this research and directions for future work are discussed in the next chapter, Chapter 6.

5.1 Ideation

In the discussion of the first iteration evaluation we determined that for SUD treatment specifically the prototype lacks a more robust way for the EPU to provide gestural cues, that the current way of perceiving the physical behaviour of the IVRU is sub-optimal, and that for use with real clients it might be better to change to EPU avatar appearance. Not specifically SUD treatment related, we determined that the prototype's ability to let the EPU experience social presence was lacking and that measures to keep better track of the whereabouts of both user's avatars for the EPU could be beneficial as well. Aspects of the prototype that were indicated to not require any changes were the number and variety of actions provided on the dashboard, the available awareness cues, and the appearance of the IVRU avatar.

The resulting requirements that complement the requirements that were established in the first iteration are presented in Table 6. Note that the origin of these requirements are the first iteration evaluation (FIE) and that the user requirement with UR ID 2 concerns the ability of both users to communicate to establish common goal and strategies, exchange what is perceived, and discuss assessed progress. This requirement was indicated to possibly be hindered by the EPU avatar's appearance.

The requirements that are likely to have the most impact on the prototype are those that relate to the way the EPU uses the desktop to interact (R22, R23, and R24), as changes to this fundamental aspect might resonate in the software-side of the setup as well.

5.1.1 Idea generation and prototype concept

We held a brainstorm to generate ideas on how to accommodate the aforementioned requirements. Sketches of the resulting 15 ideas are shown in Figure 28. Most ideas focus on specific hardware additions to enhance the EPU experience rather than changing how the treatment provider interacts with the conventional desktop setup.

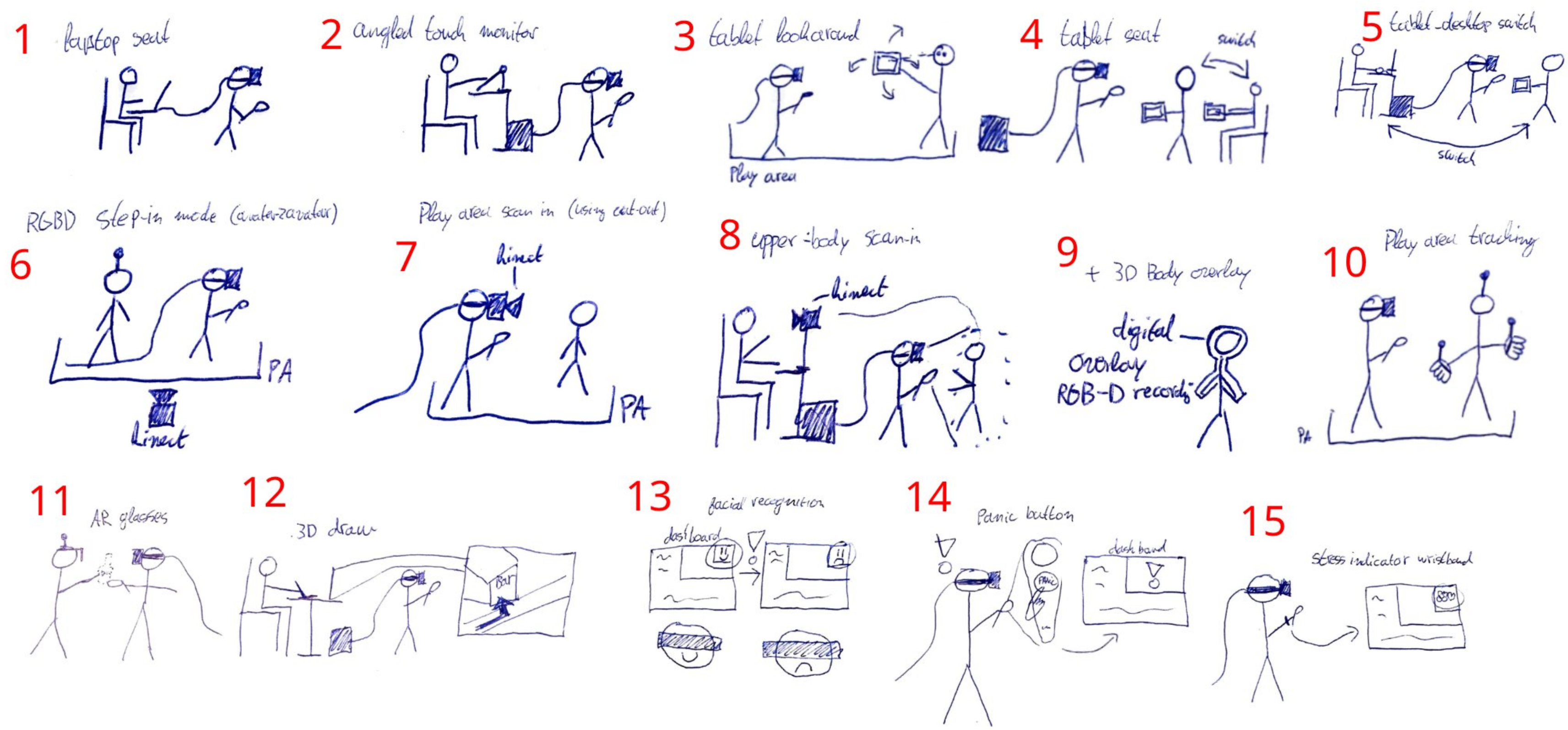


Figure 28: 15 sketches for ideas on accommodating newly identified requirements. Each idea is titled as follows: laptop seat (1), angled touch monitor (2), tablet look-around (3), tablet seat switch (4), tablet-desktop switch (5), depth image step-in mode (6), play area scan-in (7), upper-body scan-in (8) + avatar 3D body overlay (9), play area tracking (10), AR glasses (11), 3D draw (12), mood recognition (13), panic button (14), and stress sensing wristband (15).

Table 6: Extension of the requirements presented in Table 2 for an AVR solution to support joint exploration for SUD treatment for people with MBID

ID	Type	Requirement	Priority	Origin	UR ID
22	UR	The treatment provider can perceive and act on the physical behaviour of the client.	Must	FIE	-
23		The treatment provider can perform gestures to convey emotional reflection and validation.		FIE	-
24		The treatment provider can feel as if being together with the client in the VE	Should	FIE	-
25		The treatment provider can keep track of the whereabouts of their representation and the representation of the client.		FIE	-
26	FR	The solution presents information on the physical behaviour of the client to the treatment provider or allows the treatment provider to perceive the physical behaviour of the client directly.	Must	FIE	22
27		The treatment provider can provide gestural cues in the VE based on dynamic input.		FIE	23
28		The treatment provider can observe hints as to where their and the client's representation are whenever these representations are obstructed by other elements in the VE or are outside the view of the treatment provider.	Should	FIE	25
29	NFR	The solution allows the treatment provider to feel together with the client while interacting in the VE.	Should	FIE	24
30		The treatment provider's virtual appearance represents a normal human being.	Could	FIE	2

Instead of picking three ideas with the most potential as we did in the first iteration, we think a combination of two particular ideas seems to cover the additional requirements, whilst being limited in adding complexity to the AVR setup and use thereof. These two ideas are: (i) presenting the EPU dashboard on a portable touch tablet and (ii) allowing the EPU inside the IVR play area and use IVR tracking technology as an additional way of interacting with the IVRU.

By making the dashboard portable by means of a touch tablet, the EPU can adjust their position in or outside the IVR play area to always be able to properly perceive the physical behaviour of the IVRU. In addition, the portable display makes it possible to still perceive the VE as well while using the IVR tracker technology to provide more elaborate gestures in the VE. Being in close proximity to the IVRU and having the possibility to directly interact with them might increase the feeling of togetherness for the EPU as opposed to looking at the interaction from an exocentric perspective. When inside the play area, a representation of the EPU needs to be visible for the IVRU as well to avoid collisions while users move within the boundaries of the play area. Having a visible representation of the EPU can also help with the requirement of keeping track of the representations as the relative position between representations is the same as that of the relative position between users while inside the play area, although this still needs to be compensated for when the EPU is outside of the play area. Additionally, this setup makes it easier for the EPU to physically intervene when the IVRU moves on a collision course, as was considered a concern by one of the participants of the first iteration evaluation given they were behind a desk. A sketch of the second iteration prototype concept is provided in Figure 29.

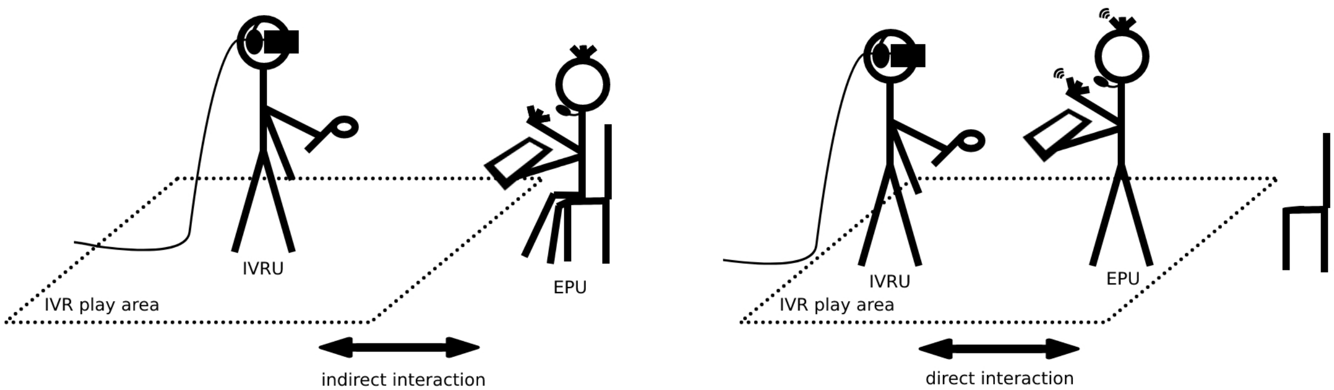


Figure 29: Sketch of the second iteration prototype concept. On the left, the EPU sits next to the IVR play area and interacts with the IVRU via the dashboard presented on a touch tablet. On the right, the EPU stepped inside the play area and makes use of touch tablet and wearable IVR trackers to interact with the IVRU. The EPU can switch between the two ways of interaction by stepping in and outside of the IVR play area.

Note that we think that other ideas with different variations of the identified properties could be used to cover the requirements as well. For example, we could retain the static desk position of the EPU and focus on providing the EPU with additional awareness cues about the client's physical, or even physiological, behaviour on the dashboard, think of a way to provide elaborate hand gestures by means of (3D) image capture, and provide the EPU means to control an egocentric perspective of their avatar such that the IVRU's interaction is directed more to them. However, it seems more sensible to leverage the availability of the IVR technology and its intuitiveness of use. By limiting the technological means necessary to accommodate the requirements, we can more easily avoid making the solution too technically complex for practical use.

As the interaction inside the VE does not seem to substantially change, we did not deem it necessary to sketch out another interaction scenario given the new prototype concept. Hence, we continue with the description of how the concept was realised in the second iteration prototype.

5.2 Second iteration prototype

The following paragraphs describe the changes and additions compared to the first iteration prototype. Elements of the first iteration prototype that are not discussed remain the same as described in Chapter 4.

5.2.1 AVR setup components

An updated visualization of the main and sub-components for the second iteration prototype is provided in Figure 30. For the EPU, the monitor and mouse are replaced by a touch tablet, headwear, and a glove with all three having an IVR tracker attached. The position of the tablet, head of the EPU and hand of the EPU, can be determined inside the IVR play area with the use of trackers. To provide the dashboard on the tablet and receive touch input back, the desktop application is streamed to the tablet using Unity Render Streaming[12] using a peer-to-peer connection via a browser. To setup this peer-to-peer connection, a signaling server is required to run on the desktop PC. The microphone component of the EPU and IVR set component of the IVRU did not change. Compared to the conventional setup, this setup might make it even easier for the EPU to remain aware of their real surroundings (R21) due to the portable display. Having clear what components the AVR setup consists of and how these components are used by the EPU and IVRU, the interactions these components enable interaction in the VE are as follows.

5.2.2 Tracked and untracked mode

To easily distinguish between the two modes of operation, we refer to them as *tracked mode* (TM) and *untracked mode* (UM), where TM refers to the mode where the EPU interacts inside the IVR play area using the IVR trackers and the UM to where the EPU remains outside of the play area where the IVR trackers have no effect. We first describe the UM and how it compares to the interaction of the first iteration prototype after which we describe how to switch to TM and what interaction possibilities that mode provides.

Untracked mode

In UM the interaction is similar to the first iteration prototype, however, the EPU uses touch controls on the touch tablet instead of the mouse for input. The touch gestures that were deemed recognizable and logical to use for their corresponding actions were chosen from the reference guide by Villamor *et al.* [94]: tap to press a dashboard button or select an object or location in the VE via the main interaction window, tap again to perform the chosen action using the selection, double tap to quickly select and perform, drag with a single finger to move the point of focus, two finger rotation to rotate around the POF, and two finger pinch and spread to adjust the camera's perspective. Note that the camera control modes for varying levels of following the IVRU avatar still apply. Overall, the available actions on the dashboard are the same with the addition of a button next to the camera control menu to switch to TM. The control hints in the main interaction window and icons in the camera control

[12] About Unity Render Streaming: `https://docs.unity3d.com/Packages/com.unity.renderstreaming@3.1/manual/index.html` [visited 27-April-2022].

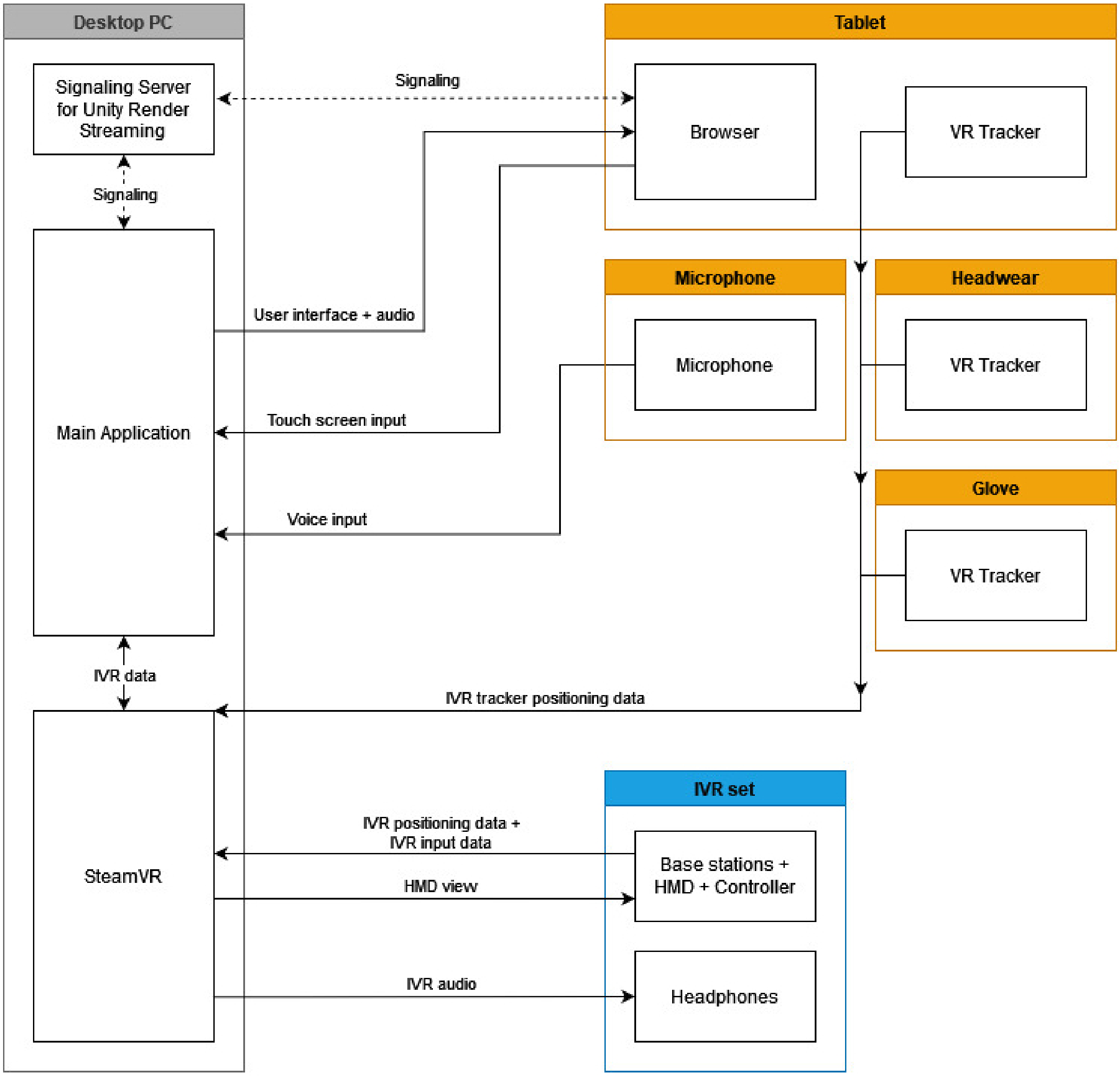


Figure 30: Main and sub-components of the second prototype and how they interact with each other. Compared to the first iteration prototype, the desktop peripherals component is replaced by four separate components, marked orange, that are used by the EPU: Tablet, Microphone, Headwear, and Glove. To provide the dashboard on the tablet via the browser Unity Render Streaming is used, which requires a signaling server to run on the desktop PC to set up the peer-to-peer connection between browser and Unity application. The IVR component, marked blue, is used by the IVR and did not change. The black arrows between the components indicate the flow and what type of data is transferred between those components. The dotted black arrows indicate that the signaling is temporary until a peer-to-peer connection is established.

Figure 31: On the left, control hints for the camera touch controls as shown in the main interaction window. From top to bottom: drag with a single finger to move the point of focus, two finger rotation to rotate around the POF, and two finger pinch and spread to adjust the camera's perspective. On the right the camera control menu with updated icons and the button to switch between tracked and untracked mode.

menu are updated accordingly, as shown in Figure 31, which also shows the button to switch between modes. The way of interacting via the dashboard remained practically similar to the first iteration prototype and as such the accommodation of requirements in that regard should still be uphold in the second iteration prototype.

What was added to the dashboards main window are additional awareness cues regarding the position of the EPU and IVRU's avatar within the VE. These awareness cues regard the avatars being obstructed by other elements in the VE or being outside the view of the EPU (R28). Whenever an avatar is outside of the EPU's view, a coloured arrow with an image of the avatar points towards the location of the avatar at the edge of the main interaction window. The arrow and avatar image are coloured blue for the IVRU avatar and orange for the EPU avatar. If the avatar is within the view of the EPU but barely visible due to the height of the camera, the avatar image is displayed on the location of the EPU avatar. Whenever an avatar is obstructed by other elements in the VE, the avatar becomes brightly coloured and outlined which are visible through the obstructing elements. Examples of the awareness cues are shown in Figure 32. Note that requirements that regard changing the EPU avatar's appearance to represent a normal human being (R30) and helping the EPU associate with the avatar (R31) did not make it into this iteration's prototype. While note concerning avatar appearance, tracked mode can help with the EPU associating with the avatar as well by essentially becoming the avatar, similar to the IVRU.

Figure 32: Examples of obstruction (left) and out-of-view (right) awareness cues for the avatars as presented in the main interaction window of the EPU.

Tracked mode

At any moment during use, the EPU can step inside the IVR play area and switch to TM using the aforementioned switch-modes button. In this mode, three IVR trackers are used to represent the physical behaviour of the EPU via their avatar. These trackers are located on the EPU's head, EPU's hand, and back of the touch tablet. We added the switch-modes button in contrast to switching modes automatically to make sure that the EPU is in control whether they are represented via TM (R14) and, for example, not accidentally be represented whenever they have to cross a part of the play area.

Upon switching modes, the EPU avatar is placed automatically on the corresponding location relative to the IVRU using the IVR tracker located on the head. This placement process shows the same awareness cues as before using the cylinder and accompanying sounds.

In TM, the EPU can perceive the VE from an egocentric perspective on the tablet. The view adapts based on the position of the tablet tracker. In this perspective, the EPU cannot use the avatar actions or camera controls, but they can still perceive the IVRU WYSIWIS window and use the environment actions. The environment actions remained unchanged and include outlining objects, the attention grabber, and placing temporary landmarks. Additionally, the EPU can let their avatar display more elaborate gestural cues using the tracker on their hand and letting the avatar follow those movements (R27). The avatar only follows the position of the hand and does not represent the individual fingers, because we wanted the EPU to still be able to use the tablet's touch controls, which would not be possible if they were using the available HTC Vive controller.

For awareness cues on body orientation and viewing direction, the EPU avatar's body is always located at the horizontal position of the head tracker, the EPU avatar's head adjusts to the position of the head tracker, and the EPU avatar is always rotated towards the tablet tracker. Having the EPU avatar look towards where the EPU is looking via the tablet might seem more logical, because that is the EPU's viewpoint in the VE. However, not letting the avatar's head depend on the tablet position allows the EPU to lower the tablet to get a better look of the physical behaviour of the IVRU in real life (R26), without the avatar looking down and appear to be looking somewhere else than towards the IVRU. Using this setup, the EPU is required to always have the tablet positioned in front of them to have the EPU avatar represent their rotation correctly. To ensure that both avatars have the correct height with respect to their head movements, the height has to be manually calibrated before use.

Having the EPU avatar represent the actual position and behaviour of the EPU in the TM, the interaction from the IVRU is now directly aimed towards the EPU themselves. This, in addition to a closer physical proximity, might help the EPU feel more together in the VE (R29).

Having an understanding of the adjustments and additions to the prototype, the next step is to evaluate this second iteration prototype in a formal evaluation to see how it affects the interaction between users.

5.3 Evaluation second iteration prototype

The purpose of this evaluation is to gain insight on how the second iteration prototype affects the interaction within the context of SUD treatment for people with MBID to ultimately get a more definitive answer to RQ1. Hence, this evaluation needs to address the relevant dimensions within RQ1: joint exploration, collaboration in VEs, AVR for SUD treatment, presence including social presence.

This evaluation of the second iteration prototype was approved by the ethical committee of Electrical Engineering Mathematics and Computer Science of the University of Twente (RP 2021-243) and the Wetenschappelijke Commissie at Tactus Verslavingszorg.

5.3.1 Participants

Similar to the first iteration evaluation, there is an EPU and IVRU role to fulfill, with the EPU requiring to be treatment providers experienced with treating people with MBID for SUD. We decided to still refrain from involving people with MBID until the interaction experience in both roles is deemed sufficient by the treatment providers. From the first iteration evaluation we determined that interaction with the IVRU role from the perspective of the EPU was not enough to determine how that interaction would be experienced by people with MBID. Treatment providers having expertise on the behaviour of people with MBID with regard to SUD, we decided to fill the role of IVRU with a treatment provider as well, resulting in participant pairs.

5.3.2 Approach

To ensure that in the second iteration evaluation participants are provided experiences regarding all aforementioned dimensions, we established constraints for each dimensions that should be accommodated towards during the evaluation. The constraints and their accompanying dimensions are listed in Table 7. Note that these constraints are based on the established requirements, as the requirements should capture the needs for joint exploration using AVR for SUD treatment.

In the first iteration evaluation we decided on an approach that involved the navigation and utilization aspect of joint exploration simultaneously. We did this in order to more easily determine large bottlenecks in the interaction provided by the first iteration prototype, before distinguishing the needs and implications of use for SUD treatment for either aspect separately. Having addressed the identified bottlenecks in the design of the second iteration prototype, we can focus on distinguishing those needs and implications in the aforementioned dimensions. Therefore, we decided to have participants perform two exercises; one more focused on navigation and the other more focused on utilization.

Because the think-aloud method was indicated to be cognitively taxing to a substantial degree in the first iteration evaluation, we ask participants to only think aloud whenever the prototype or their interactions with the prototype subverted their expectations or whenever they were struggling to express their intentions via the system. This should provide insight on where the prototype falls short without the overhead of including every aspect that is supported well.

There are various ways to accommodate the aforementioned constraints. Most importantly, there should be incentive for interpersonal communication without the need for substantial prior instructions, as was done in the first iteration evaluation. We believe that this communication is best incentivized by having a discrepancy in available information or actions for each role where both are needed to successfully complete the exercises. This should also prevent a leader-follower dynamic to emerge where there is no need for one person to act as the other can do everything by themselves.

Additionally, we decided on dedicating participants to a single role of role of either EPU or IVRU. Having a single role, they should become more knowledgeable on that role and are able to compare their experiences between exercises with different purposes.

Similar to the first iteration evaluation, by discussing the experiences of participants we can get a deeper understanding of those experiences. Hence, a semi-structured focus group is held after the exercises have been completed. Moreover, participants should be instructed on the interaction possibilities the prototype provides before the exercises and are not allowed to ask questions on how to approach certain situations given the interaction possibilities to avoid bias, also similar to the first iteration evaluation.

Table 7: Second iteration evaluation constraints per dimension and related requirement

Dimension	**Constraint**	**Related requirement**
Joint exploration	Incentive to execute the full exploration process	R1
	Incentive and the freedom to communicate to establish common goals and strategies, exchange what is perceived, and discuss assessed progress	R2
	Incentive to exchange strong synchronous wayfinding cues	R9
Collaboration in VEs	Incentive to use independently controlled views of the VE	R8
	Incentive to converge to a common visual or spatial location	R10
AVR for SUD treatment	Incentive to perceive and act on the physical behaviour of the IVRU	R22/R26
	Incentive to perform more elaborate gestures in the VE	R27
	Incentive for participants to discuss how they experience the VE and how it affects them	R2
Presence and social presence	Incentive for the participants to think about place illusion and plausibility illusion	R5
	Incentive for the participants to think about social presence	R24
	Incentive for the EPU to interact in both the tracked and untracked mode	R29
Second iteration prototype	Incentive for the IVRU to interact using the IVR equipment	-
	Incentive for the EPU to interact in both the tracked and untracked mode	-

Figure 33: Picture of the three IVR trackers used in the second iteration evaluation, attached to a tablet, hairband, and velcro hand strap.

5.3.3 Equipment and layout

The experiment was done using the same equipment and in the same room as the first iteration evaluation, using a similar equipment layout as well. Additional equipment includes a touch tablet, wearable IVR trackers, and a high-speed wireless router. Notably, the position of the EPU was changed to a chair near the edge of the IVR play area.

The IVR trackers for the head and back of the hand were made wearable by attaching a tracker to a hairband and velcro strap respectively, as shown in Figure 33. The high-speed wireless router streamed the dashboard to the tablet via a local network. Note that the monitor and mouse were used by the researcher to observe the VE during the sessions and perform required manual actions. A full overview of the equipment used, layout, and pictures of the setup is shown in Figure 34.

5.3.4 Procedure

In the procedure, two participants play out two exercises with a short break after each exercise. After the exercises, their experiences during the exercises are discusses in a focus group.

Introduction
Similar to the first iteration evaluation, the participant in the IVRU role is instructed to act like a general client with borderline intellectual functioning and an alcohol use disorder during both exercises. For example, the persona of the

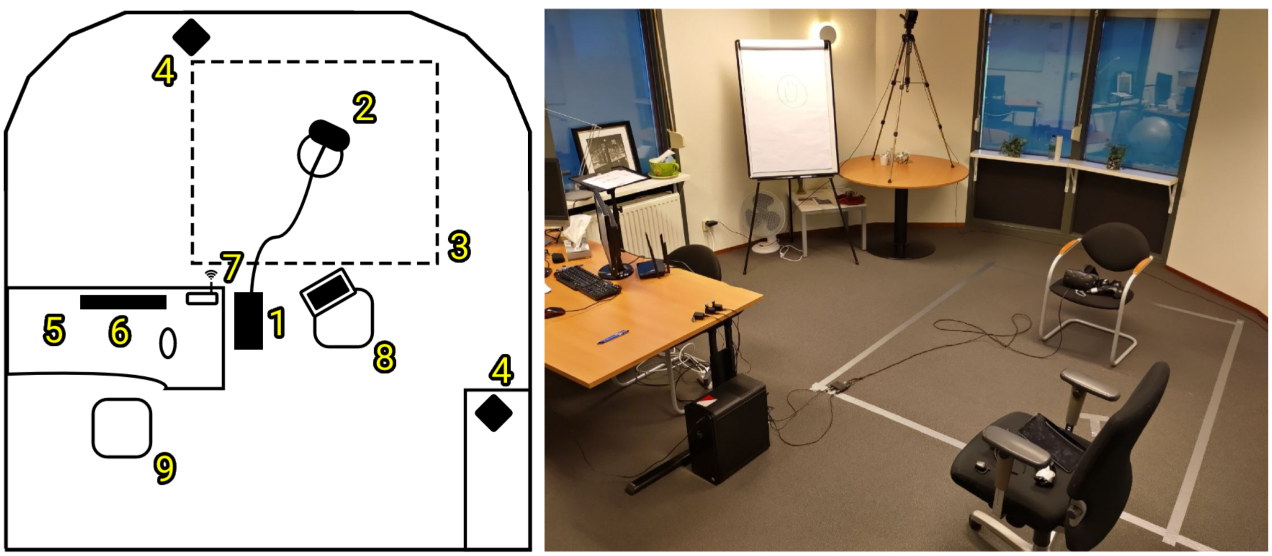


Figure 34: On the left a graphical overview of all equipment used for the second iteration evaluation and their layout: desktop PC (1), IVR HMD and sound dampening headphones for the IVRU (2), outline of the IVR play area (3), IVR base stations (4), desk (5), monitor and mouse (6), high-speed wireless router (7), chair, tablet, IVR trackers, and wireless microphone for the EPU (8), and chair for the researcher (9).

general client can represent the first real client that comes to mind. Additionally, the client persona is to be knowledgeable on SUD treatment practices and cannot actively disrupt the flow of the exercise. For example, being too overwhelmed by the alcohol-related cues or not wanting to use the IVR technology would be considered a disruption of the flow. In similar fashion, if at any point the provided exercise would be considered too weird or too difficult for clients and thus the client persona, the IVRU is asked to play along as if it would not be too weird or difficult. Of course, any personal reason outside of the client persona are valid to temporarily stop or discontinue the exercises. The EPU participant is instructed to behave like they normally would as treatment providers.

After the instructions specific to the client persona, the participants get instructions on how to interact via the AVR solution in their respective roles and are shown all available actions at least once. While the actions for the EPU are shown, the IVRU wears the HMD to observe how these actions appear from within the VE. This allows them to see what kind of instructions they can expect from the EPU such that they are able to refer back to specific actions during the exercises. For example, if the EPU forgets that they have the temporary landmark available, the IVRU can ask the EPU to *use the flag*, naturally reminding the EPU of that action if that action would be deemed useful.

Lastly, the EPU is instructed to keep an eye on the IVRU during the exercises and intercept if the IVRU seems to be leaving the IVR play area or is going to hurt themselves due to them being oblivious of their real surroundings. The researcher will look out for the safety of participants as well, but this instruction provides additional incentive for the EPU to observe the physical behaviour of the IVRU.

What follows are two exercises that are separated by a short break. The first exercise is more focused on the navigational aspect of exploration, whereas the second exercise is more focused on the utilization aspect of exploration. Whenever the participants are provided instructions for an exercise, the instructions are provided to the EPU who is asked to relay those instructions to the IVRU using their avatar in the VE. Providing instructions in the VE is a core part of joint exploration for SUD treatment and this helps to gain insight on that aspect specifically. Both exercises take place in the same VE, which is mostly similar to the virtual city of the first iteration evaluation, but with slight changes in building layout to accommodate the exercises.

First exercise

The first exercise is focused on the navigational aspect of exploration and the use of the untracked mode of the prototype solution. Hence, the EPU is instructed to start out in the untracked mode to provide incentive to use the untracked mode at least briefly, but they are free to change modes whenever they see fit during the exercise.

At the start of the first exercise, the participants are instructed to take a look around the VE together and discuss anything of interest, SUD related or

Figure 35: A visualization of perspectives of the EPU on the left and the IVRU on the right during the second part of the first exercise. The EPU can see the red marker pointing towards the destination, but not the roadblocks. The IVRU can see the roadblock but not the red marker.

not. This should provide participants incentive to execute the exploration process, discuss how they experience the VE, use independently controlled views, and converge to common visual or spatial locations. After five minutes of exploration, the participants are instructed to go to a specific bar within a time limit of ten minutes.

The EPU gets the location of the bar presented on their dashboard via a red marker, whereas the IVRU is provided no information on the location. What the IVRU does see is that suddenly various roadblocks appear throughout the VE which they cannot travel through. These roadblocks are not visible by the EPU, including the IVRU WYSIWIS window, and their presence is not included in the instruction. Having this discrepancy in knowledge, the participants have to communicate regarding the common goal and strategies to reach that goal, using strong synchronous wayfinding cues to convey that information. A visualization of both perspectives during this part of the exercise is shown in Figure 35.

There is a bar in roughly every quadrant of the city and the participants start the exercise roughly in the center of those four quadrants. To avoid the participants sticking around the bar too much during the initial five minutes of exploration, the bar appears to be closed. To provide participants a similar challenge, independent of where they end up during the initial five minutes of exploration, the bar that they have to navigate to is located in the diagonally opposite quadrant of the city of where they are after the initial five minutes. Upon reaching the destination or passing the time limit of ten minutes, the participants take a short break, temporarily taking off the HMD and wearable IVR trackers. If the participants did not reach their destination, the researcher manually places the participants at the destination during the break. Figure 36 shows the four quadrants of the city, the starting point of the exercise, and where each bar is located as well as the zones that are blocked off for travel.

During the break, both participants are instructed to individually fill in a short questionnaire about their experiences regarding presence and social presence for the completed exercise. After the break, the participants continue on with the second exercise of the experiment. Details on this questionnaire are provided after the description of the second exercise.

Second exercise

The second exercise is focused on the utilization aspect of exploration and the use of the tracked mode. Hence, the EPU is instructed to start out in the tracked mode to ensure that the mode is at least used briefly, but they are free to change modes whenever they see fit during the exercise. Similar to the first part, the second part of the exercise consists of two segments as well.

At the start of the first segment it is explained that there are eleven *mystery bottles* scattered throughout the bar. The content of a bottle can be revealed by having the EPU touch the bottle using the IVR tracker on their hand. The participants are instructed to look for the bottles, place them onto the bar, and determine what drink affects the IVRU the most during this process. While both users have their individual viewpoint and can move around separately within the IVR play area, only the IVRU can move the position of the play area in the VE and transport the bottles using their controller. This discrepancy in locomotion possibilities in the tracked mode require the users to coordinate their movements to complete the first segment. Again, providing participants incentive to execute the exploration process, discuss how they experience the VE, use independently controlled views, and converge to common visual or spatial locations. After finding all bottles and determining the liquor type most affecting or after passing a time limit of ten minutes, the exercise continues to the second segment. All four bars in the VE have the same layout, bottle locations, and bottle appearance as shown in Figure 37.

After a drink has been selected by either the participants or research upon passing the time limit, the participants are instructed to role-play buying a drink at the bar and discussing the role-play afterwards in the VE. During the rolelay the EPU plays a pushy bartender that tries to sell the IVRU their selected drink and the IVRU plays themselves as client trying to cope with the situation and to leave with a non-alcoholic alternative. Upon passing the time limit of five minutes or the exercises deemed completed by the researcher, the participants take a short break, again filling in the short questionnaire a second time. After the break, the participants discuss their experiences with the researcher in a semi-structured focus group.

Short questionnaire and focus group questions

During the short breaks, both participants fill in a short questionnaire about their experiences regarding the aspects of place illusion, plausibility illusion, and social presence for each of the two exercises separately. We recognized

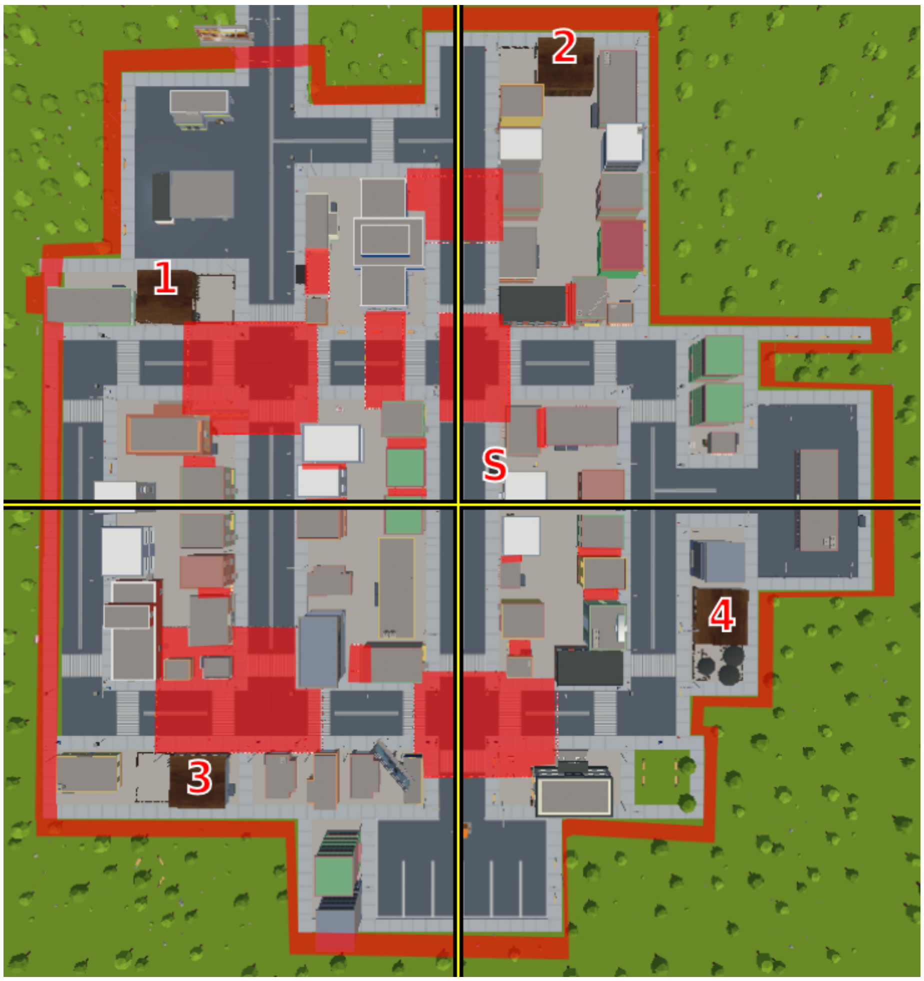


Figure 36: Top down view of the city and its separation into four quadrants. The location of the four bars are indicated by a number and the starting point of the exercise is indicated by the letter *S*. The blocked off non-travel zones are indicated in red.

Figure 37: On the left a top down view of the bar with the location of each mystery bottle in the bar numbered. On the top right, a mystery bottle as they appear in the bar. On the bottom right, a mystery bottle identified by having the EPU touch the bottle using the IVR tracker on their hand. The contents of the bottles in the bar are: Baileys (1), wine (2), whiskey (3), cocktail (4), Dutch gin (5), vodka (6), beer (7), rum (8), cider (9), digestif (10), and Amaretto (11).

Table 8: Second iteration evaluation questionnaire questions

Left end side	**Right end side**
I was not aware of my real surroundings	I was very aware of my real surroundings
I had the feeling that the other and I were active on different locations	I had the feeling that the other and I were active on the same location
I did not understand the intentions of the other at all	I did understand the intentions of the other clearly
The visual communication of the other was very unclear	the visual communication of the other was very clear

that experiences on these concepts might mostly be subconscious and perhaps hard to recall after the fact. Therefore, these questionnaires are not meant for obtaining quantitative data on these concepts, rather they function as an incentive for participants to actively think about and form an opinion on these concepts before being asked about them during the focus group.

Questions of the short questionnaire and related questions of the semi-structured focus group are based on the igroup Presence Questionnaire [44] and Networked Minds Social Presence Questionnaire [7] and were provided to the participants in Dutch using the provided Dutch translation for igroup Presence Questionnaire questions and self-translations for Networked Minds Social Presence Questionnaire questions.

The four questionnaire questions are presented on a visual analog scale with the first statement presented on the far left and the second statement presented on the far right and markers indicating the middle and quarters of the line. The participants were instructed to mark a point on the line that represents their opinion on the statements most. The statements for each side of the scale for the questions are presented in Table 8.

For the semi-structured focus group, we established five main questions regarding the aforementioned dimensions that we focus on in this iteration's evaluation. Each main question was accompanied by related follow-up questions to stimulate the focus group if necessary. A translation of the focus group questions can be found in Table 9. The original Dutch version of the short questionnaire and focus group questions can be found in Appendix F.

5.3.5 Processing and analysis

The focus group results are processed in similar fashion to the first iteration interviews using inductive content analysis, including coding, grouping, and finding structure. The top level codes were established *a priori* and correspond to the five dimensions that this evaluation focuses on.

Table 9: Translated second iteration evaluation semi-structured focus group questions

Dimension	**Translated focus group questions and optional follow-up questions (marked by a ∼)**
Joint exploration (navigation) + collaboration in VEs	How did navigating through the city together go? ∼To what extent could you understand the intentions of the other? ∼Were there elements of the system that sat in the way of navigating together? ∼Were there interaction possibilities that you were lacking?
Joint exploration (utilization) + collaboration in VEs	How did gathering the bottles and the role-play go? ∼To what extent could you understand the intentions of the other? ∼Were there elements of the system that sat in the way of gathering together? ∼Were there interaction possibilities that you were lacking?
Second iteration prototype	How did you experience interaction via the tablet and position sensors? ∼Which way of interacting had your preference and why?
Presence and social presence	Did you have the feeling that you were taking on the virtual world together and what made you feel that way? ∼Did you have the feeling that you were together in the same virtual world and what made you feel that way? ∼To what extent were you aware of the other in the virtual world? ∼To what extent did you have the feeling that you were paying attention to each other? ∼Did you have the feeling that you were interacting inside the virtual world or did you have the feeling that you were controlling something from the outside and what made you feel that way?
AVR for SUD treatment	How do you envision the use of such a system for exercises with real clients? ∼Where do you forsee problems with interaction via the system for therapy exercises? ∼How aware were you of the real environment?

5.4 Results

5.4.1 Participants

The experiment was performed with three pairs consisting of one male and five female participants between 24 and 38 years old (Mean: 30.5, SD: 5.128). The participants will be referred to as participant 1 through participant 6 or P1 through P6 for short. All odd numbered participants (P1, P3, and P5) fulfilled the EPU role, whereas all even numbered participants (P2, P4, and P6) fulfilled the IVRU role. Of the participants, participant 1 and participant 2 were the same participant 1 and 2 from the first evaluation. The other participants did not partake in the first evaluation.

Of the participants, only participant 1, 2, and 3 had prior, although limited, IVR experience. Participant 1 already had IVR experience before the first evaluation, participant 2 obtained more IVR experience in the time between the two evaluations, and participant 3 had previously tried out observing virtual scenes using her smartphone and a cardboard viewer.

5.4.2 Results

The results of the second iteration evaluation are separated into four main topics: *joint exploration and collaboration in VEs*, *presence and social presence*, *AVR for SUD treatment*, and *prototype and procedure*. The prototype and procedure topic regards comments of the participant specifically on the second iteration prototype or the procedure of the experiment. For readability, the results are further grouped into related sub-topics. The full coding structure is provided in Appendix G. Full transcriptions, including codes, are available to other researchers upon request by contacting the corresponding supervisors of this research.

Joint exploration and collaboration in VEs

Awareness cues

In tracked mode as IVRU, participant 2 indicated that she did not really focus on, or notice, the awareness cues produced by the EPU avatar, but she indicated that the hand gestures supported the *realness* of the interaction. In response, participant 1 indicated that she was constantly looking down at the tablet, which resulted in a lack of eye contact displayed by the EPU avatar. However, participant 2 indicated that she did not notice the lack of eye contact due to being busy in the VE.

Participant 3 and 4 indicated no problems with the awareness cues produced by the avatars. They do note that the hollow eyes of the EPU avatar were intense "it was only that, like I said, with the eyes that I thought 'Woah'" (P4). They add that being able to look along directly with the IVRU was nice and made it immediately clear what was referred to during certain interactions.

Communicating intentions

Participant 1 had the urge to communicate mainly verbally, because sequentially performing additional actions was experienced to be too high of a cognitive load in the untracked mode. Participant 1, 2, 3, and 4 specifically, and participant 5 and 6 more indirectly, indicated that the verbal communication was most important to have, compared to the other gestural and graphical options. Participant 3 and 4 specifically added that the exercises could not have been completed without verbal communication, and participant 5 and 6 noted that doing everything verbally "went fine".

Regarding the communication with the EPU, participant 2 indicated that she mainly focused on perceiving the verbal communication of the EPU, because the EPU avatar's gestural communication was experienced as off-putting and sometimes confusing in both tracked and untracked mode "those movements she made with her arms and stuff, those were not really... those are mainly weird movements which made you think 'What are you doing?'". This in contrast to participant 4 and 6, who indicated that in the tracked mode the intentions of the EPU were more clear when the EPU avatar mimicked the hand movements of the EPU in contrast to the untracked mode where those hand movements were deemed less clear. Participant 6 added that the EPU avatar making hand gestures that fit logically with the corresponding verbal communication made the depiction *more complete*, because "I think that you are doing things more literally".

Participant 4 noted that the options used to convey intentions were often clear, but the true meaning of the intention was sometimes difficult to understand, especially in the first exercise. However, he indicated that that might be because of being overwhelmed by the possibilities of IVR technology.

For going to the bar in the first exercise, participant 3, 4, 5, and 6 found that simply walking ahead with the EPU avatar to show the way worked best for communicating possible routes, where participant 6 specifically noted that the avatar stayed in view and the EPU kept asking whether she could follow, which was sufficient for the exercise. Participant 3 noted that the EPU avatar did not always behaved as expected, due to the automated pathing and for the EPU invisible blockades, making it especially confusing to ensure that the correct intentions were communicated in the untracked mode.

More concretely for the first exercise, participant 1 and 2 indicated that roughly 80% of the time they would communicate solely verbally, and in the remaining 20% they would use the additional cues in situations where verbal communication only does not suffice. Participant 5 indicated that being able to walk with the EPU avatar and communicate verbally suffices and that the lack of use of other options was not experienced as a problem on her end.

Furthermore, participant 4 noted that the temporary landmark was difficult to see due to it being a large pole with a similar appearance to the street lights at eye height. Participant 1 also noted that in the tracked mode, she wanted to perform an applause by pressing the button, but realised only then that she had to do that herself, which she could not due to having the tablet in hand.

Interaction possibilities of the EPU, relative position of users in the IVR play

area, and autonomy of the EPU
Participant 1, 3, and 5 all indicated a lack of being able to directly interact with the VE themselves as EPU. For example, grabbing and moving items like the IVRU was able to with their controller. Participant 1 specifically indicated that this was due to a lack of not feeling present in the VE, which caused a general feeling of not being able to do anything when using the tracked mode, despite being able to look and walk around the IVR play area for herself. Participant 5 added that the technology that should have enabled more natural interaction made the interaction more unnatural "the tablet in your hand and sensor on your hand are not natural". Participant 4 added that he noticed that when the EPU used the tracked mode, the EPU could do less than he could, however, this did not impede the collaboration from his end. Participant 3 simply expected to be able to do the same as the IVRU in the tracked mode, but without the need for an HMD. Additionally, participant 3 expected to still be able to use the avatar actions that are normally available in the UM, however, she indicated that that was due to a lack of experience.

According to participant 3, the small field of view of the tablet in the tracked mode made it hard to get an overview of the VE and what the client was doing, which was much easier in the untracked mode. She added that being able to watch along with the IVRU in the tracked mode helped in that regard, a view that participant 1 shared, however, she indicates that being able to still get an overview position in the tracked mode would be nice. On the other hand, participant 3 noted that having this overview position might result in losing a certain connection to the IVRU in the VE. She indicated that shifting that connection to the IVRU in real life might restore that connection again. Participant 5 added to this by mentioning that in the for her it was sometimes already hard to see what the IVRU was referring to in the untracked mode. For participant 5, standing still when using the tracked mode made it much easier to keep an overview and it was also much clearer what she could do with the technology. Additionally, she mentioned that the tablet sat in the way of more naturally interacting with the IVRU, however, the tablet was currently a necessity to remain aware of the VE. This view was somewhat shared by participant 1, who indicated that in both the tracked and untracked mode, you had to remain focused on the tablet.

In the tracked mode, the relative position of users in the VE remains the same as the relative position of the users in the IVR play area, also when teleporting. Participant 1 and 3 indicated that this required a new mindset during the exercise, where participant 3 indicated that this was not weird, but participant 1 indicated that this new mindset was difficult and as such a substantial cognitive burden. Participant 5 noted that this retention of relative position was outright confusing and made her lose orientation "For me it was difficult. I was finally standing somewhere, I lose my orientation, I am somewhere else".

Participant 1 and 2 would have preferred a larger IVR play area for the second exercise with no teleportation at all, as the hassle with teleportation and relative position took away from the *realness* of the VE. Participant 4 had a similar opinion, but called out teleportation in both tracked and untracked mode for breaking the spatial illusion a bit when using IVR. Participant 5 indicated that at the end of the exercise it went a bit better if she remained close to the IVRU, which did not necessarily induce a feeling of not being able to move, but rather it did induce a feeling of lack of control or autonomy, which was indicated to have much implications on being a treatment provider "Because I was trying to make things work quite a bit and how can I be a therapist then?".

Presence and social presence

Presence experienced by the EPU

Participant 1, 3, and 5 did not experience place illusion in the VE during either exercise. Participant 1 and 5 specifically indicated that their lack of feeling present in the VE was not a problem, as long as it is uphold for the IVRU, but participant 1 added that it might have been nice to have. For the second exercise, participant 1 indicated to find it important to feel present, which was currently not the case. She adds that the lack of feeling present in the VE in the second exercise for finding the bottles originated from a lack of means to interact with the VE herself, such as grabbing objects or teleporting around, compared to the IVRU and that the experience of presence was higher during role-play, but still not fully.

Participant 3 indicated that she would feel less present in the VE if she knows she is represented by the EPU avatar's current appearance. Participant 5 indicated to just mostly being busy with an *outside screen*, which resulted in communication towards the IVRU as a person in the real world, rather than towards the IVRU avatar in the VE.

Presence experienced by the IVRU

Participant 2 indicated to feel less presence in the first exercise, whereas participant 4 indicated to feel more aware of his real surroundings, without explicitly mentioning the notion of presence, and that he did not experience this as a problem. Both participant 2 and 4 noted that the sound dampening headphones did help her feel more present and less aware of his surroundings respectively.

Participant 2 explained that in the first exercise, when the EPU avatar was not in view because it was not there or walking behind her, she would have to speak to the real world, diminishing presence. She indicated that if the avatar would have kept walking next to her during teleportation and always be visible in the corner of her eyes it would have helped to maintain the feeling of being inside the VE. Similarly, in the second exercise, when she made a physical connection with the EPU, this was experienced as a connection with the physical world and as such hindering presence "it makes you go back to the real world". She noted that this is solely for physical connection with other people and that mimicking inanimate objects, such as walls, with real objects would make the VE more real. Aside from the physical connection, during interaction with the EPU in the second exercise when the tracked mode was used, she felt much more present "Then I really forget what I am doing here".

Participant 4 indicated that having to teleport diminished the *realness* of the experience overall and would have preferred a larger play area and only being able to walk inside that area. Additionally, during the second exercise, participant 4 indicated to be much less aware of his surroundings. In contrast, participant 6 indicated to feel fully present in the VE in both exercises.

Social presence experienced by the IVRU

During the first exercise, participant 2 and 6 indicated to be mainly unaware of the presence of the EPU avatar. They noted this because when the avatar was out of view, for example when automatically following behind them, the EPU avatar was completely silent if the EPU did not say anything. Participant 6 indicated that because of this, the EPU avatar's appearance would sometimes startle her and at the start of the exercise she would actively avoid the EPU avatar. Participant 2 added that having the avatar walk next to her when following would have made a difference, but for her in the first exercise the EPU avatar did not really have added value. In contrast, participant 4 indicated to always be very aware of the EPU avatar in both exercises. Participant 2 does note that the EPU's voice coming from the EPU avatar helped her with "being in that environment".

In the second exercise, participant 2, 4, and 6 all indicate an increased feeling of positive *togetherness* with the EPU. Participant 2 and 4 specifically noted that the tracked gestures made it "more real", where participant 4 specifically added that the EPU could have been in another room and it would still have been more real. Participant 2 added that the EPU being able to interact somewhat more with the VE, such as touching the bottles and standing behind a counter, made it look more real, but on the other hand she noticed that the EPU had different capabilities than her in the VE "That is a bit weird, because it does not fully add up". On the other hand, participant 4 indicated that when the EPU was in the tracked mode, he had the feeling that the EPU could do more than in the untracked mode. Participant 6 notes that in the second exercise she was simply more aware that the EPU avatar would always be in her near vicinity, which was strengthened by doing the exercise together, which in turn increased a feeling of togetherness. Participant 4 also indicated an increased feeling of satisfaction accompanying the increased feeling of togetherness.

Participant 2, 4, and 6 all mentioned that a more realistic EPU avatar appearance would be beneficial. Participant 2 noted that making the EPU avatar appear more real would help make the environment and interaction more real as well, whereas participant 4 and 6 argued the other way around and that the current appearance make the EPU avatar appear more fake and in the case of participant 4 even less personal. Concretely for participant 4, with a more realistic EPU avatar appearance "then you have the feeling that you are really talking with someone".

On the topic of physical interactions between users, participant 2 and 6 indicated that having a physical interaction in the VE made the EPU avatar more fake, where in contrast participant 4 indicated that it made the avatar more real. Participant 2 suggested that this comes from an increased connection with real life, where participant 6 simply described it as a "mind fuck" where feeling a real hand, which is definitely not real in the VE, was experienced as disturbing. Participant 6 indicated that having an improved EPU avatar appearance might also improve the physical touch experience with the avatar. Participant 4 argued that a more formal touch, such as shaking a hand, showed him and the IVRU, "Hey, this is really me". He added that having such a physical connection "just feels better". Participant 2 noted that physical contact with clients is exceptional and can arouse fear easily, especially with traumatized clients and especially when blindfolded by the HMD. Participant 5 added that from her EPU perspective, the physical connection did not add "something extra" regarding the interaction.

Representation of the EPU and IVRU via their avatars
Participant 1 explicitly indicated that she was someone outside the VE that were using the EPU avatar to provide guidance inside the VE. She also had the urge to keep checking how she was represented in the eyes of the IVRU via the designated WYSIWIS window. This urge was indicated to originate from a "lack of trust" in the system to represent her properly. For example, "I wanted to check if I was looking at her and that everything I do is done properly". Participant 1 and 2 both acknowledged that having the three point tracking might be insufficient to ensure that the EPU feels well represented inside the VE. They argued that having the avatar display as much mimicry as possible would eventually get the most realistic interaction experience, which includes knowing you are well represented and as such free up the cognitive space now used for ensuring this.

Participant 3 and 4 indicated that the EPU avatar behaviour in the untracked mode felt much more "robotic" and in the tracked mode "much more personal", the latter indicated to be nicer to interact with. Participant 4 added that the voice coming from the avatar really helped making a connection "I really had the feeling that when I was looking at the avatar that I was making contact with [name P3]". Participant 6 indicated to have "zero, zero connection with the orange figure" and that having participant 5's voice coming from it made it only slightly better. She added that you can quickly lose out of sight that a trustworthy someone is behind the avatar with its current appearance. Participant 2 and 6 both indicated that having the avatar look like a normal person would help much with the presentation and making the interaction more real. Participant 6 added that for her it would make the avatar more trustworthy as well and it "would make it easier to visit worlds together", as at the very start of the exercise she indicated to be actively avoiding the EPU avatar. The avatar's appearance aside, participant 2 and 6 recognized that the EPU avatar is a point of contact within the VE that allows you to speak to the EPU inside the VE. As participant 2 pointed out, when the avatar is not there, you speak towards the physical world.

Lastly regarding the EPU avatar, participant 6 noted that, aside from the EPU's voice, the avatar was completely silent "I compare it to an electric vehicle", which could easily startle you. During the first exercise, the EPU avatar kept walking in front of her and the EPU kept asking her whether she was still there and that was sufficient for the exercise. In the tracked mode, having the EPU always in close proximity to the IVRU helped participant 6 to keep an overview of the whereabouts of the avatar.

Regarding the representation of the IVRU via their avatar, participant 3 and 5 indicated that the IVRU avatar was there, but they would mostly interact with the real IVRU. Participant 3 clarified that she was more concerned with the blue figure inside the VE, despite contact with the actual IVRU and participant 5 indicated that it felt "more logical" for her to direct her communication towards the real IVRU, because she was not inside the VE.

General behaviour of the EPU avatar

Due to a bug in the setup, participant 1, 2, 3, and 4 noticed that the EPU avatar was constantly slowly moving towards the IVRU, which was an uncomfortable experience. Participant 4 specifically described it as an invasion of his personal space. This bug was fixed for the session of participant 5 and 6 to not further disturb remaining participants, hence they could not notice it. While teleporting or placing the avatar, participant 3 and 4 mentioned that the avatar sometimes appeared uncomfortably close to the IVRU when using the untracked mode. In tracked mode, participant 3 and 4 noted that the sizeable appearance of the avatar made the EPU appear closer than they might have stood in real life, despite being calibrated to have the same height as the EPU. Participant 1 said that the avatar simply needs to behave otherwise it can be very disturbing, especially for traumatized clients. However, she added that the avatar's behaviour would likely not be blamed on the treatment provider as it will "remain a game" for clients.

During the first exercise with the roadblocks that were invisible for the EPU, both participant 3 and 5 indicated that the avatar was sometimes suddenly gone because it took an unexpected route, which was very confusing for them. Participant 3 added that while walking, the avatar sometimes also took unexpected routes, such as through alleyways, where she would have liked to follow the main road.

Regarding letting the EPU avatar follow the IVRU, participant 2 indicated that having the avatar walk behind her made her forget about it and it would have been more logical if the avatar would walk next to her. Lastly, participant 5 tried to let the avatar look at a building, which was currently not possible, as it always looked towards the IVRU.

Being together in the VE

Participant 1, 2, 3, and 4 all indicated that they had the feeling their attention was directed towards each other. Participant 1 and 2 noted that this feeling was strengthened by noticing that the other reacts to you verbally and visually. For participant 3, the tracked mode helped in that regard "I think that is because you are going to look together and go for it" and being able to point at something which the IVRU was able to observe and having that shared interaction helped as well. For participant 4 as IVRU, whenever the EPU was using the untracked mode, the verbal communication via the headphones really induced the feeling of “we are going to take a walk together”.

Participant 3 noted that she prefers to really work together with a client and that sitting behind a desktop, watching along, and only occasionally giving instructions would have created a feeling of separation as opposed to togetherness. Lastly, participant 5 indicated that she simply was aware that she was constantly present for the IVRU.

AVR for SUD treatment

Awareness of real surroundings

Participant 1 and 3 both indicated that they were mostly focused on the tablet while using the tracked mode, resulting in a lack of awareness of their real surroundings, occasionally remembering that they had to keep an eye on the IVRU as well to help them avoid collisions. Participant 1, with support of participant 2, indicated that this was due to a high cognitive load from the technology and exercise combined in either mode, “you have to take way too much stuff into account at once”. She added that with more experience she would probably be able to check out the surroundings as normal.

Participant 3 indicated that for her it was the lack of overview on the VE that the tracked mode provided that constantly kept her focused on looking around using the tablet. She added that she would be subconsciously more aware of her surroundings when using the untracked mode with the overview perspective, especially if she would be more familiar with the system or the layout of the VE.

For participant 5 it was the other way around. She indicated to be decently aware of her real surroundings while using the tracked mode, but less so in untracked mode. She indicated that this was because she struggled to get the hang of the tracked mode and as such was mostly “out of it” and therefore had the opportunity to be aware. She adds that would she have been more “in it”, including her role as treatment provider, she would probably be less aware.

Observing physical behaviour of the IVRU

For participant 1, 2, 3, and 4, the face was indicated as the most important part of the body to be able to observe as treatment provider, however, the face of the IVRU was fully covered by the combination of HMD and hygienic face mask. They indicated that most of the time feelings could probably be communicated verbally. Participant 4 with support from participant 3 indicated specifically that the face being covered was “a loss, but I don’t know if it is a large one”.

Observing the posture was deemed less important than the face by participant 1 and 2, but they recognize that treatment providers who specialize in body language might argue otherwise. Participant 3 and 4 recognized the importance of observing body language, however, participant 3 noted that it might be barely visible. Participant 4 added that he was sometimes startled during the exercise, which should have been visible in his body language, but participant 3 indicated not to have noticed it, mostly due to being too busy with getting an overview of the VE via the tablet. Additionally, participant 1 and 2 recognized that due to the use of IVR technology the postural behaviour of clients might already differ substantially than compared to similar situations in real life.

Participant 5 indicated that, while using the untracked mode, she would be looking at the IVRU during conversation. She indicated that this was due to a lack of feeling present in the VE "because I was not in that world I think", making it more logical for her to focus her conversation towards the IVRU in real life. She added that in both modes, she could not observe bodily reactions due to struggling with getting control over the system, which she indicated to possibly improve with time.

To the idea of having a wristband that measures heart rate and skin conductivity to present the therapist with a measure of stress, participant 1 noted that it might help identifying when clients experience stress. However, they noted that it is yet another thing to take into account in an already cognitively burdening system.

Use with real clients

Participant 1 and 2 mentioned that the current level of mimicry the EPU avatar displays is probably sufficient to be interpreted correctly by most clients, but it does depend on the cognitive level of the clients. Additionally, participant 2 noted that the extra effort that is required by the treatment provider to get to this level of mimicry is worth it compared to the benefits it brings for the client. Participant 6 indicated that when the client knows the treatment provider is with them in the IVR play area, it would simply feel nicer for her. Participant 5 added that during use with real clients, it is important that at least the treatment provider is in control, which was not always the case during the exercises.

Participant 2, 4, and 6 as IVRUs all indicated that the VE itself can be distracting for some clients, hindering the actual exercise. Participant 2 noted that this distraction would probably mainly occur with clients with lower cognitive ability. Participant 4 indicated that a few practice session might overcome this distraction, as clients can get easily impressed by the IVR technology and thus become distracted. Participant 6 argued that it might also have to do with the size of the VE, as a larger environment is more likely to trigger a natural curiosity that distracts from the exercise, whereas in the relatively small bar environment this natural curiosity would be less "You are less inclined to look around. That simply was the setting and you move in that setting". Participant 4 added that distractions from real life does not help with easily distracted clients as their focus will shift to their real surroundings. However, wearing the headphones did help in regard to blocking outside sounds.

Moreover on the EPU avatar, participant 3 indicated that the EPU avatar with the voice of the treatment provider might be confusing for clients "I can imagine that it brings confusion like 'Hey, this is not you, but I can hear you'". Participant 3 added that making the avatar more realistic also might trigger a more realistic response from the clients. Participant 5 and 6 indicated that when the client just looks around the VE by themselves, the EPU avatar might even be redundant to have, however, they are uncertain how confusing just hearing the treatment provider's voice inside their head can be for clients.

Before they would want to use this setup with real clients, participant 3 and 4 indicated that some of the "clumsiness" that comes with the setup needs to be fixed, such as accidentally teleporting inside a building with no seemingly way out, as that can arouse fear and stress for clients and can easily sit in the way of the goal of an exercise. Participant 5 and 6 indicated that there is a need for a well-grounded client-treatment provider relation to make effective use of such a system, because without knowing a treatment provider well, being put in IVR as a client can feel "very threatening". Lastly, participant 5 and 6 also indicated that if something is not visually clear for the treatment provider and they cannot adjust their view in such a way that it becomes visually clear in the untracked mode, they think it would be possible to instruct a client to stand or look slightly different to get the desired view with the current system and that intrinsic changes are not required for the system in that regard.

Prototype and procedure

Habituation

Participant 3 and 5 both explicitly indicated that once more habituated they might be able to use the system better. Participant 3 added that normally, when going outside with a client, she tries to have a conversation. However, during the exercise she was too busy trying to figure out everything and being impressed by the system's capabilities. Participant 5 also indicated that for a first time use in such an experiment it was much to take in "I couldn't use all the things I tried out, because I simply didn't have the time". She added that it might get better with time, but she doubted whether it would come to her easily.

Interpretation exercises

Participant 1, 3, and 4 indicated that they expected exercises to be more therapy related given the treatment provider and pseudo-client role, where participant 1 mentioned that the role-play exercise came closest. Participant 2 indicated that the bottle search exercise was nice to start with, however, immediately seeing what is inside the bottle would result in cravings with real clients and as such hinder the rest of the exercise if the exercise was given in a therapeutic context.

For the navigation exercise with roadblocks specifically, participant 1 thought that not being able to see the blockades was a fault in the system, which frustrated participant 2 as she kept noting that the blockades were really there. Participant 5 also noted that for her it was not clear that only the IVRU could see the blockades. Notably participant 6 did not inform participant 5 about the barrier's existence, which resulted in unexpected behaviour of the EPU avatar, making the exercise confusing for participant 5.

For the second exercise, participant 5 and 6 noted that the exercise on its own already helped to create a feeling of togetherness, not just the EPU using the tracked mode. Participant 2 indicated that teleporting while having to keep a grip on the bottle was hard, especially because the bottles would fall out of her grip if they bumped into something. Losing the bottle when bumping into something was also noted by participant 4. Participant 1 and 2 explicitly indicated a preference for using the tracked mode in the second exercise, however, participant 1 noted that for the first exercise, using the untracked mode was sufficient as well.

Interpretation experiment procedure
Participant 2 noted that the system introduction of the EPU happened after she was already wearing the HMD, so she did not know how the EPU used the system to get to the visible result on her end. Therefore, she could not picture the EPU using the system during the experiment. Additionally, she added that when speaking too loudly, it sometimes resulted in an echo where first the real voice would be heard and thereafter the voice via the microphone in the VE.

For participant 3 and 4, it was not clear that it was the task of the EPU to explain the exercise to the IVRU. Because participant 3 did not convey the exercises, participant 4 got the feeling that he was missing important information. Additionally, because everything was a new experience, the subject during conversation was often the new technology, making participant 4 forget his role as client as he mentioned. Lastly, participant 5 indicated that getting to know everything about the system and the exercises in a relatively short time was essentially too much to absorb it all in one go.

Interpretation second iteration prototype
Participant 2 indicated that the noise dampening headphones really helped with staying focused on the VE. Participant 4 noted that the grid wall, that became visible when nearing the edges of the IVR play area, and the red indication when teleportation was not possible, were clear. He added that he would have liked to feel a bit of haptic feedback from the controller when bumping with his avatar into the environment. Participant 5 indicated that it was not clear for her what the difference in interaction possibilities was between the tracked and untracked mode. Finally, participant 6 explicitly indicates that she thinks that this system can already help with triggering real reactions.

5.5 Discussion

The purpose of the second iteration evaluation was to gain insight on how the second iteration prototype affects communication to ultimately get a more definitive answer to RQ1: *How to facilitate communication for joint exploration to support SUD treatment of people with MBID using asymmetric VR?*. We set up an experiment that focused on five dimensions that we identified regarding RQ1: joint exploration, collaboration in VEs, AVR for SUD treatment, presence and social presence, and novelties of the second iteration prototype. We discuss the results of the second iteration evaluation for each dimension separately, as well as noteworthy miscellaneous results that fall outside these dimensions, and limitations of this evaluation.

5.5.1 Joint exploration

We aimed to provide participants incentive to execute the full exploration process and communicate using strong synchronous wayfinding cues. Communication between users was to be on establishing common goals and strategies, exchange what is perceived, and discussing assessed progress.

Exploration process

We expected participants to equally contribute to each other's exploration process in both exercises, however, our results show that this was not the case. We noticed that every EPU let the IVRU take the lead at the start of the first exercise, followed by the EPU taking the lead in the second part of the exercise due to knowing their common destination. In both cases there was seemingly equal opportunity to share thoughts and information to come to common goals and strategies. For example, the EPU could have highlighted the destination so that that information is presented to the IVRU as well, but in most cases they did not.

Notably, this contribution was more equal in the second exercise, seemingly suggesting that the interaction in the first exercise was influenced by either one or both of two things: (i) the EPU's perspective was deemed superior by the EPU for the second part of the first exercise, as it seemed to provide the destination and full survey knowledge, or (ii) there is a specific social dynamic for exploration tasks where the client takes the lead when there is no clear goal and the treatment provider takes the lead when there is a clear goal. We think the interactions that we observed across sessions was influenced by both, given that in the bottle search in the second exercise the perspectives were more equal and the goal was clear for both users.

Communication on exploration

The use of verbal communication between participants persisted throughout both exercises with gestural and graphical cues only used occasionally if necessary, often to support verbal messages. A notable exception to this was the EPU using their avatar to lead the way in the first exercise to provide the IVRU with route knowledge, while using verbal comments to support this visual action.

Communication on how the VE affected the IVRU was mainly done verbally, with only the IVRU sometimes using gestural cues to refer to specific elements of the VE. Communication on spatial tasks, such as navigating to the bar in the first exercise and bottle searching in the second, was in contrast more accompanied by visual cues produced by both EPU and IVRU among all sessions. Notably, during the bottle search in TM, the EPU seemed to be slightly more generous with gestural cues as well.

When and how often particular cues are used seems to be a balance between the ease with which the cue is produced and the necessity of the cue for the exploration process. For example, the EPU participants seemed to reserve their use of visual cues when using the tablet in untracked mode for when they were deemed necessary, whereas in the bottle search brief pointing gestures were made more often when using the IVR tracker in the tracked mode, although in the search these gestures seemed to be less necessary. Participant 1 specifically indicated that the actions in the UM required more cognitive effort due to their sequential input method. The additional less useful gestures to verbal communication were noted to make the communication more complete by participant 6, indicating expectations about communication which the EPU avatar in the UM was not able to accommodate.

Participant 4 and 6 did add that the gestures in the TM did help communicating intentions, whereas in UM the gestures did less so. For participant 2 this was not the case for either mode. She did not explicitly indicated a distaste for gestural cues, but rather the way they were currently presented by the EPU avatar.

Overall, communication for exploration seems to be accommodated on a practical level at least for similar tasks. There seemed to be a general preference for communicating via the TM, however, UM was found to be more practically useful for spatial navigation tasks, such as the first exercise where the overview position was deemed useful. How this communication was experienced by both users via their avatars is discussed in its respective dimension: presence and social presence.

5.5.2 Collaboration in VEs

An aspect of collaboration that was noted by every EPU which we did not account for, is that they had wrong expectations about their interaction possibilities upon using tracked mode in the IVR play area. Specifically, wrong expectations about not being able to directly interact with the VE themselves, for example teleporting or grabbing items, similar to the IVRU. We specifically left out the use of a controller by the EPU in tracked mode to still allow for dashboard access. However, the dashboard remained practically unused in tracked mode across all sessions. From the perspective of the IVRU, this difference seems to be not as impactful as it was only brought up once by participant 4 where it was deemed not to be problematic for the collaboration. Knowing this, incorporation of a controller and levelling the playing field further is a change that can be easily incorporated.

While using tracked mode, the IVRU has partial control over the viewpoint of the EPU, as the IVRU controlled the location of the IVR play area by means of teleportation. This phenomenon was indicated by all three EPU participants of which participant 1 and 5 specifically mentioned that it hindered the joint experience. For tasks in relatively small VEs, the solution suggested by participant 1 and 2 of using an IVR play area that covers the entirety of the VE is a valid suggestion. However, this limits the size of the VE to the available space in real life, which can be undesirable as well.

Awareness cues

Participants seemed to be content with the current cues. However, we did not realise that most awareness cues produced by the EPU avatar were solely visual, with the exception of the bell sound during placement. Participant 2 and 6 noted that, for example, the sound of footsteps might have helped to keep track of the avatar when it is outside the IVRU's field of view, which often occurred while using the follow action.

Additionally, avatar behaviour that we did not account for was that when the EPU held the tablet in a more horizontal position, for example to look at the IVRU in real life directly, looking down to check the tablet made the avatar look down towards seemingly nothing for the IVRU, as the tablet was not visualized in the VE. While this did not seem to hinder the interaction, it might let the EPU avatar come over as distracted or malfunctioning. Visualizing the tablet on the other hand can get in the way when holding the tablet in front of the face to directly observe the VE, making the avatar look at the tablet instead of the IVRU, again looking distracted.

5.5.3 Presence and social presence

We found that none of the EPU participants experienced place illusion throughout the exercises. This can be seen as a positive, as the EPU needs to ensure the safety of the client and ensure the integrity of the nearby environment in real life. Feeling present on another location might hinder that. However, if that safety and integrity can be ensured in a different way, the EPU experiencing place illusion was indicated to possibly have a positive effect on the joint exploration experience from the perspective of the EPU. Inexperience with the prototype supposedly resulted in the EPU having to focus mainly on the tablet display, leaving limited room to inspect the nearby environment at the same time.

On the other hand, experiencing place illusion for the IVRU seemed to be the baseline. Specific occurrences could diminish that experience, such as directing communication towards real life instead of the VE or the unnatural way of locomotion. Teleportation was identified to potentially diminish place illusion at the cost of avoiding VR sickness before [29, 50], but since we only recorded this occurrence once, we think the choice remains valid. On directing communication towards real life, implementing one of the earlier discussed initiation techniques, such as the magic flower, might help to keep the communication within the VE.

Togetherness

All IVRU participants indicated an increased feeling of positive togetherness with the EPU while they were using the tracked mode. For the EPUs, only participant 1 and 3 explicitly indicated that they felt more together while using the tracked mode as well. The feeling of togetherness was indicated to be caused by how both users react differently, both visually and verbally, on each other's behaviour. In contrast, interaction in untracked mode was identified to be more artificial by participant 3 and 4. The difference between TM and UM regarding the feeling of togetherness suggests that there is more to gain with embodied interaction within the SUD treatment context.

The EPU avatar appears to be able to relate to the IVRU in either mode, helping the plausibility illusion to occur. However, the perceived realism of the EPU avatar in the tracked mode seems to help bolster this occurrence. We recognize that in the exercises, the user avatars were the only elements of the VE that could address the other directly and therefore be the only source of plausibility illusion for the users. We deem the chances for plausibility illusion to occur for the EPU low in either mode, as in TM they would be directly addressed by the IVRU in real life and in UM it is not them who is being addressed by the IVRU, but their avatar of which they felt to be a distant pilot.

Social presence
Regarding social presence, the EPU avatar appearance was noted to be of importance, specifically for SUD treatment. However, we did not realise how much the current EPU avatar appearance would affect communication outside of the practical aspect, to the point where participant 6 would actively avoid the avatar in the first exercise. Having the EPU avatar take on a more human-like appearance might provide a more pleasant experience with the avatar.

One occurrence regarding social presence that triggered a different reaction in every IVRU participant was a physical connection with the EPU while they were in TM. Participant 2 lost spatial presence while participant 4 gained spatial presence, but participant 6 became anxious from the experience. The reaction of participant 6 was indicated to be due to making something real that should clearly not be real, which seemingly refers to the test-dummy like appearance, and as such might be avoidable by having the EPU avatar appear more human-like and friendly, possibly looking like the EPU themselves.

5.5.4 Interaction for SUD treatment

Perceiving physical behaviour of the IVRU was still deemed important by participants, similar to the results of the first iteration evaluation. Participants of the second iteration evaluation indicated that for them it was most important to observe the face of the IVRU, which was currently covered fully by the HMD and an hygienic face mask. While the reasoning behind the EPU having a dynamic position still applies to better perceive the IVRU from any desired angle, providing information about facial expressions when wearing an HMD is substantially more complex. Providing additional information on this matter on the dashboard, in contrast to perceiving the physical behaviour in real life, might provide opportunities here. We do recognize the comment of participant 1 and 2 that while using IVR the physical behaviour of clients might be substantially different compared to real life, but we are currently unable to confirm that.

While Participant 2 and 4 specifically mentioned that having the EPUs hand gestures tracked provided an interaction that felt *more real*. Unfortunately, the results provide limited insight on how these tracked hand gestures can be used to better convey emotional reflection and validation. Participant 4 indicated that while the gestures were practically useful, their true intention was often unclear. Additionally, we already identified that in the current prototype, more physical gestures, such as a handshake or pat on the shoulder, can impact the experience substantially either positively or negatively. Moreover, the comment of participant 5 that having a tablet and sensor on your hand is not natural, similar to clients possibly changing their behaviour when using IVR technology, might refrain EPUs from providing gestures to convey emotional reflection and validation themselves.

Treatment provider in control

While conducting similar exercises, the treatment provider wants to be in control of the system. Throughout the sessions it was clear that some habituation is required to fully get a grip on the solution and its capabilities. This might hinder participants forming an actual opinion about certain aspects of the system instead of ascribing it to inexperience. Therefore it could be beneficial for future evaluations to have a try-out session a day or week before the actual experiment in which participants have time to get more accustomed to the solution, so that their focus can lie on intended use and how they experience that.

On a related note, a treatment provider always wants to be in control over the system when conducting exercises with clients to ensure that the client is provided a suitable experience, which the current inexperience of EPUs clearly contradicts. While we identified elements of the system that the treatment provider has less control over, such as the automated pathfinding of the EPU avatar, what we did not realise was part of this control was trust in the system representing the EPU correctly. From the exocentric perspective in untracke mode, the EPU could practically always see the behaviour of the EPU avatar, but they could not check the avatar's behaviour while using tracked mode. Participant 1 made this explicit by indicating doubts about the avatar representing her physical behaviour correctly by only wearing the three trackers, making her constantly check the IVRU WYSIWIS window to check her avatar's behaviour during communication. While she was the only EPU with this concern, it does suggest there is value in letting treatment providers experience the client's IVRU perspective to ensure that the experience is suitable.

Additional quirks in the current system that were ought to be resolved before use with real clients were: having a more human-like EPU avatar, ensuring that the EPU avatar cannot violate the personal space of the IVRU, having the EPU avatar display more audible awareness cues, and that the IVRU cannot accidentally teleport inside places they are not supposed to be.

5.5.5 Interaction using the AVR solution

There seemed to be a desire for an experience that is considered a joint experience from the perspective of both roles, which currently is less so for the EPU in UM. Despite being more physically together with the desk removed from the setup, using UM seemed to provide less satisfaction in regard to having a joint experience, whilst being practically useful.

On a more technical note, the noise dampening headphones were indicated to help with place illusion, but sometimes the EPU could still be heard through the headphones when they were talking too loud, resulting in an echo. Notably, wearing the headphones made the IVRUs speak louder which was sometimes picked up by the Bluetooth microphone as well, which also resulted in an echo. Since both users are to remain co-located, only upgrading the used technology seems to help circumvent both problems.

5.5.6 Limitations

Having three sessions with three pairs did provide the insight of six treatment providers. We dedicated one role to each participant, so that they would become essentially experts in that role given that they experienced both exercises from that perspective and could provide more in-depth information on those experiences. While this argumentation might still uphold, it does substantially limit the number of opinions per role. Two alternative approaches are swapping roles after the first exercise and having the EPU use both the UM and TM in each exercise or having an additional session with slightly different exercises at a later point, effectively doubling the number of participants for each role.

While we attempted to cover both the navigation and utilization aspect of exploration, this did seem to sit in the way of the exercises being representative for SUD treatment, as was indicated by participant 1, 3, and 4. Role-play was indicated to be the most representative, which only was a five minute segment at t he end of the second exercise. This did provide us more information in joint exploration using AVR in general at the cost of getting more focused information of joint exploration using AVR for SUD treatment.

Having discussed the implications of the second iteration evaluation results on the five dimensions relevant for RQ1, we passed the narrow middle section of the hourglass as presented in the introduction and can continue by discussing the implications of the results obtained in both evaluations on the main research questions and joint exploration using AVR in general.

6 Discussion and conclusion

Over two user-centered design iterations we built an asymmetric virtual reality (AVR) solution supporting joint exploration for substance use disorder (SUD) treatment of people with MBID in an attempt to answer the research question of this study and speculate on joint exploration using asymmetric virtual reality in general:

> **RQ1: How to facilitate communication for joint exploration to support SUD treatment of people with MBID using asymmetric VR?**

With the accompanying sub-questions:

> **SQ1: What are critical aspects that an AVR solution needs to adhere to to make that solution suitable for use in SUD treatment?**
>
> **SQ2: What is necessary for effective communication for joint navigation?**
>
> **SQ3: What is necessary for effective communication for collaboration in virtual environments?**

In this chapter we discuss the implications of this study on SQ1-3 and RQ1 as well as this study's limitations and possibilities for future work. We start by aiming to provide more definitive answers on SQ1-3 in Chapter 6.1-6.3 respectively. Implications of this study that are relevant for providing a more definitive answer on RQ1 that are not yet covered by SQ1-3 are discussed in Chapter 6.4. The attempt at answering RQ1 in Chapter 6.4 is followed by discussing the limitations of this study in Chapter 6.6 and possibilities for future work in Chapter 6.7. Finally, this work is concluded upon in Chapter 6.8.

6.1 Critical aspects AVR for SUD treatment

When we established an initial design rationale, we identified the cruciality of clients, *i.e.* the immersive virtual reality user (IVRU), obtaining and maintaining presence as well as producing ecologically valid responses during immersive virtual reality (IVR) experiences. Our results affirm the importance of both aspects.

Next to the aspects identified in the initial design rationale, our evaluations provided additional insight on what treatment providers believe to be crucial for such AVR solutions. We identified the following four additional aspects throughout both evaluations. (i) being able to communicate verbally and (ii) being able to watch along the perspective of the client, both with regard to creating a mutual understanding of the IVR experience between client and treatment provider. (iii) Having a virtual representation of the treatment provider should never display inappropriate behaviour nor violate the social boundaries of the client and (iv) the solution should not require physical interaction between users, both with regard to their potential to make the client uncomfortable, especially considering traumatized clients, resulting in an unpleasant interaction that possibly leads to discontinuation of the exercise.

To a certain extent, these findings are shared by Skeva *et al.* [81], who performed a qualitative analysis on the considerations of IVR for SUD treatment based on expert opinion, which was not yet available at the time of establishing requirements of the first iteration. They highlight the need for eliciting ecologically valid responses as well, but the role of presence is not explicitly mentioned. (i) and (ii) are explicitly mentioned with regard to creating a mutual understanding of the experience as well. Regarding (iii) and (iv), while they do not go into inappropriateness of avatar behaviour or physical interaction between users specifically, they do highlight the risk of traumatization or retraumatization of users given certain IVR experiences.

Notably, these aspects, except for (iv), were to certain extent already covered by the requirements on joint exploration for the first iteration prototype. (i) was covered by the requirement for exchanging strong synchronous wayfinding cues. (ii) is a specific implementation of the requirement of the treatment provider always being able to understand and act on the perspective of the client. And for (iii) we took into account the social conventions of traditions and imposed a 1.5 meters social boundary around the IVRU and have the avatar display predictable standardized behaviour.

Although (iv) was not explicitly considered during either iteration, our generated ideas do not reflect any requirement of, or potential for, physical interaction between users. This does not exclude the possibility of incorporating physical interaction for joint exploration using AVR. For example, Gugenheimer *et al.*'s FaceDisplay [38] as described in Chapter 2.4 provides the externally partaking user (EPU) insight on the perspective of the IVRU and interaction via strong synchronous cues via the displays attached to the HMD and not wearing headphones. However, the solution requires the EPU to come relatively close and interact with what can be considered an extension of the body of the IVRU, which they found resulted in a high level of dominance of the EPU over the IVRU.

These additional crucial aspects could have been identified earlier in the process, by for example performing a more thorough inquiry of treatment providers' needs similar to Skeva *et al.* [81]. However, we argued against such an elaborate inquiry and favoured modest field expert involvement in the requirements engineering phase and letting participants experience a realized prototype to base their opinion on first. Given the similarities in results with regard to interpersonal interaction, there does not seem to be a clear optimal approach in that regard.

6.2 Communication for joint navigation

When establishing the initial design rationale, we identified based on the navigation process as proposed by Darken and Peterson [22] that communication on joint navigation is likely to consists of establishing goals and strategies, exchange what is perceived, and discussing assessed progress. Furthermore, strong synchronous wayfinding cues are to be used for this communication and specific social dynamics between group members are to be taken into account as well as explained by Dalton *et al.* [21].

Evaluations of both UCD iterations contained exercises for which communication on joint navigation was required. Throughout these exercises we observed participants communicate on the aforementioned topics using the available wayfinding cues. All participants in both evaluations were able to complete these exercises without any sign of clear limitations to this communication, suggesting that the aspects identified in the design rationale were sufficient.

The three types of strong synchronous wayfinding cues (*i.e.* verbal, gestural, or graphical cues) are context dependent as well. Within our research, verbal communication was integral for either presenting information on joint navigation or supporting presented static visual cues. The variety of gestural and graphical cues was found to be a positive, considering the interpersonal differences between clients and how they interpret these cues. Other contexts might require a different dynamic between these three cue types or less types altogether. For example, the virtual search and rescue presented by Bacim *et al.* [4] relied mostly on static graphical cues with optional verbal cues whereas Thoravi Kumaravel *et al.* [89] relied on dynamic graphical annotations.

Notably, the social dynamic of treatment provider and client was predetermined in our research context. We did not consider this social dynamic with regard to communication, but rather with regard to accessibility of the users, where the available actions for the client were kept limited and easy to understand. This resulted in an unequal access to strong synchronous wayfinding cues in the virtual environment (VE), with both users having access to verbal and gestural cues, but only the treatment provider having access to graphical cues as well.

This inequality of cue types was not indicated to be problematic or strange by either the IVRU nor EPU. This could be because treatment providers arguably have a higher status within the social dynamic, and therefore it is not unreasonable for them to have access to additional cues. Another possibility is that users being aware of the asymmetry of technology makes an asymmetry in available actions not unreasonable for them either. However, some IVRU participants did indicate that they were mostly unaware of how the EPU used the prototype to perform actions in the VE, leaving room to question the asymmetry of actions in that regard.

If the same implementation would be used for a context where both users are considered equal, such inequality in available actions might result in an unwanted shift in the social dynamic, again noting the importance of considering the specific social dynamic that is aimed for or may arise from specific combinations of group members. In this research this became apparent in the tracked mode where only the IVRU could control movement and grab objects, resulting in a feeling of helplessness for the EPU, who desires control over the VE in their role as treatment provider.

6.3 Communication for collaboration in virtual environments

Similar to communication for navigation, we have to make a distinction between what is necessary for communication and what would be beneficial for an established solution. For the topic of collaboration in VEs in the initial design rationale, we established the need for awareness cues, consideration of social conventions of transitions, and four characteristics identified by Yang and Olson [105]: (i) users can remember or recognize the environment, (ii) users have an independently controlled view, (iii) users are able to converge to a common visual or spatial location, and (iv) users can understand each other's objective and subjective perspectives. Furthermore, we made an attempt to generalize the navigation process of Darken and Peterson [22] to apply to exploration as a whole, including the utilization aspect as well.

Starting with the four characteristics for remembering or recognizing the environment, we relied on user's own capabilities to remember spatial features of the VE, limited the number of actions available for the IVRU, and aimed to present the actions available for the EPU in a clear and concise manner via colour-coded dashboard windows. While we identified no clear limitations in participant's capabilities to remember elements of the VE, the importance of being able to remember or recognize the VE showed when participants mentioned that they might have used different actions for specific tasks if they had remembered those actions being available.

This highlights the importance of either designing interactions that are familiar for users so they are able to quickly pick up on them or prior application training before practical use. In the SUD treatment context, both design approaches apply, as intuitive interactions are beneficial for use by IVRUs with lower cognitive abilities and prior training is applicable for treatment providers regardless of design familiarity to ensure that they are in control of the application before use with clients.

EPUs having their own perspective in the VE was indicated to be a necessity for collaboration in the VE by participants who appreciated it being an option. This became more apparent when their viewpoint was being controlled by the IVRU in the untracked mode, which sometimes resulted in confusion and a loss of personal agency, both hindering the collaboration. Whenever there was less need for collaboration, participants indicated that focusing on watching along with the perspective of the IVRU was preferred to better understand the IVRU's perspective, which is in line with Stefik *et al.* [84]. From having individual viewpoints arises the need to be able to converge to a common spatial or visual location and both exercises would seemingly be substantially be hindered without such individual viewpoints.

The aspects of understanding each other's perspectives and awareness cues are closely related. An example where this shows from our solution is that in the untracked mode, the EPU avatar always looks towards the IVRU so that the EPU appears to be engaged with the IVRU with no further effort required from the EPU themselves. While throughout our evaluations this appeared mostly not to be an issue, one time the EPU was able to see a specific building from their exocentric perspective, but could not turn their avatar towards that building to make the IVRU aware that they could see the building. This resulted in the IVRU not understanding that they shared a common visual location although from a different perspective.

This example clearly shows the need for displaying awareness cues and being able to understand each other's perspectives, but also shows a flaw in our approach of disconnecting the awareness cues displayed by the EPU and their actions. Disconnection of actions and awarenes cues is seemingly not bad practice per se. The EPU avatar was acknowledged as the point of communication with the EPU and results show that people were hesitant to exclude the avatar from communication while also indicating a preference for the exocentric perspective in certain scenarios. A simple solution for specifically this problem in style of the current approach could be to have an additional *look-towards* action for the EPU to let the EPU avatar look towards a selected object or direction instead of the IVRU avatar. Do note that there is likely a limit to how many more actions can be added before it hinders application use and as such a more fundamental change in the way of interacting might be more suitable.

Lastly, we argued that necessities for the navigational aspect of exploration were applicable to cover for the utilization aspect of exploration as well by showing that the navigational process proposed by Darken and Peterson [22] could be generalized to cover both aspects of exploration. Therefore, the prototype was realized under the assumption that the necessities for the utilization aspect of exploration were covered by the necessities for the navigational aspect of exploration. We did not explicitly evaluate this generalization, however, there were seemingly no clear limitations in the prototype's ability to support the utilization aspect of the exercises of both iterations. This suggests support for our generalization within the context of this research.

6.4 Communication on joint exploration for SUD treatment of people with MBID using AVR

Having attempted to answer SQ1-3, we explored the necessities for enabling the communication for joint exploration to support SUD treatment. What remains is determining the implications of combining these three aspects within a SUD treatment context to determine how this communication should be shaped given the treatment of people with MBID using AVR for RQ1. Two main themes emerged from the information we obtained: embodied interaction and shared physical spaces. We discuss these two main themes first and thereafter we go briefly onto facilitating communication for joint exploration using AVR in general.

Embodied interaction

Embodied interaction is about using the physical body while interacting with surrounding technology, but is also about the way that physical and social phenomena unfold in corresponding environments [24]. For this research the relevant topics that stood out were: human qualities of the interaction, affective communication, and presence.

While the first iteration prototype and untracked mode were deemed sufficient for their practical purpose of facilitating communication for joint exploration, that communication undesirably lacked human qualities from the perspective of both the EPU and IVRU. In the tracked mode, the EPU was able to display more human behaviour in the VE, such as more elaborate and somewhat jittery body movements, and in turn the EPU can feel more together with the IVRU in real life. However, cueing and the human quality of communication were still hindered by the unnatural way of having to perceive the VE via the tablet, inequality of interaction possibilities for the EPU, and the inhumanity of the EPU avatar. Where addressing the inhumanity of the EPU avatar by changing its appearance is relatively straightforward, the unnatural way of perceiving the VE and inequality of interaction possibilities are more complicated due to their dependence on the hardware implementation as well.

We do not yet know how elaborate the physical behaviour of the EPU needs to be mimicked to be deemed sufficiently embodied in the VE for SUD treatment specifically. However, a possible solution that covers both perceiving the VE and leveling the interaction possibilities is having the EPU perceive the VE via a augmented or mixed reality HMD instead of the tablet. This allows for more elaborate gestures using both hands and makes the use of controllers by the EPU more viable again as they do not need to interact with a touch display anymore. Both users wearing HMDs and using controllers levels the symmetry in possible interactions between EPU and IVRU as well. Where certain AR or MR HMDs might fall short for this purpose is in how limited their field of view is. The limited FOV provided by the tablet in our prototype was explicitly indicated by one participant to hinder them in getting an overview of the VE. In general, a limited FOV can reduce performance in tasks related to spatial awareness [15], such as exploration.

For affective communication, the verbal aspect of affective communication is easily satisfied by recording voice, but providing accompanying gestures, including social touch, are less straight forward given that gestures include one's personality as well. This personality seemingly depends on appearance and specific executions of certain gestures. Ensuring that anyone's personality can be accommodated calls for dynamic appearances and input methods. For example, one person might prefer to greet with a handshake and another with a fist bump. Using IVR controllers that can track individual finger movements, such as the Valve Index controller,[13] allows for an easy way to incorporate more

[13] Valve Index Controller: `https://www.valvesoftware.com/en/index/controllers` [visited 08-May-2022]

Figure 38: Valve Index Controller finger tracker example. Still from video *Valve Index controllers handling slow finger movements* (00:16) by SadlyItsBradly from https://www.youtube.com/watch?v=cjXSXmHZP3Q [visited 31-May-2022].

elaborate gestures already while maintaining the ease of button input. An example of tracking finger movements using the Valve Index Controller is shown in Figure 38. For this topic it is important to reiterate that the use of social touch was indicated to be limited in the context of SUD treatment due to possible traumas of clients and as such the experience of touch, either physically or virtually, should be considered with care.

Having an understanding of the importance of supporting affective communication with gestures, what should not be overlooked is the information on affect that the face and psychophysiological responses can provide, in addition to perceiving body posture as was done in this research. For the EPU standing physically close the IVRU can provide some information the wellbeing of the IVRU, however, the face of the IVRU remains largely covered by the HMD and not all psychophysiological responses are perceivable via the eye. Eye tracking technology that is present in some HMDs, such as the HTC Vive Pro Eye[14] can be utilized to measure pupil dilation, which is a predictor of emotional arousal [12] and psychophysiological responses via a wrist band in order to alert the EPU of responses that they could not perceive themselves.

The problem of not being able to see the IVRU's face can be approached the other way around as well, by making the IVRU's eyes perceivable in real life. For example, Matsuda *et al.* [56] shows a way to present a three-dimensional view of the user's face and eyes in a perspective-correct manner. This approach makes the EPU having to rely less on the computer's interpretation of the facial

[14]HTC Vive Pro Eye: https://www.vive.com/us/product/vive-pro-eye/overview/ [visited 08-May-2022]

responses and might make the interaction from the EPUs perspective more pleasant as well as it enables eye contact instead of talking to the bare front of the HMD.

Getting affective communication right is for the benefit of both users understanding each other's subjective perspective on the VE, which is especially important for the SUD treatment context. This in contrast to understanding each other's objective perspective in the VE, which is important for the practical execution of collaborative tasks.

Lastly on presence of both users. We already discussed the importance of the IVRU experiencing presence for SUD treatment, however, the role of presence for the EPU seem to relate again to the distinction between practical and personal communication. For the practical angle of communication, we determined in this research that the EPU does not seem to require to experience presence.

On the other hand, for the human quality and naturalness of communication, we need to consider the EPU's use of less immersive technology. Being less immersive, the EPU is likely to remain partially aware of their real surroundings, including the IVRU, as well as the technology they are using. As we have discussed earlier with respect to embodied interaction, being aware of the technology used can take away from the naturalness, and consequently the human quality, of the interaction.

Additionally, if there is a disconnection in location between location of the EPU's avatar and the EPU themselves with respect to the IVRU, the IVRU is going to direct their communication towards the EPU avatar instead of the EPU, likely impacting on how that communication is experienced for the EPU on a personal level.

Shared physical spaces

Up until now, we mainly considered interpersonal interaction and how that is experienced in the virtual environment. Yet, for this interaction using AVR there are three additional physical spaces to consider: the general space the AVR interaction takes place in, the space the EPU has available for interaction, and the IVR play area which the IVRU has available for interaction. For the general space, we consider all space required to house the necessary equipment including both spaces for EPU and IVRU.

We saw that letting the EPU and IVRU share the IVR play area provides possibilities for embodied interaction and can result in an increased feeling of social presence for the EPU and IVRU. However, this came with substantial implications on locomotion for both users, which may hinder the AVR interaction altogether. Hence, when considering sharing the IVR play area, finding a fitting locomotion technique should be a main priority. For example, by using more elaborate multi-user teleportation techniques such as those evaluated by Weissker *et al.* [98] or substantially increasing the size of the IVR play area and forgoing locomotion outside the play area altogether.

For SUD treatment, the EPU having their separate space outside of the

IVR play area remains important. For certain exercises, the treatment provider might need to distance themselves from the IVRU in order to not show their presence when the IVRU is expected to explore the VE by themselves, with the EPU just watching along.

Purpose, size, and AVR setup complexity are of importance for the general space. Considering use in clinics, it is likely that the AVR solution will be used in one of two places: a treatment provider's office or a dedicated location within the clinic. The larger the spatial footprint of an AVR solution becomes, the less likely it fits within a treatment provider's office space and a dedicated location is required. If the AVR solution is to be set up in an office, a complex solution that is time-consuming to set up takes away from time that could have been spent on the exercise itself. For a dedicated location, confidentiality is still to be taken into account.

Considering that Tactus' programs include group exercises, such as role-play, group AVR exercises are not an unlikely next step. Not accounting for hardware limitations and logistics, aside from solving locomotion for multiple users, there seems to be no obvious hurdles that prevents the approach present in this research from including multiple IVRUs, EPUs, or both.

6.5 Communication for joint exploration using AVR in general

Having a better understanding of the implications of AVR for joint exploration in the SUD treatment context, we broaden our scope to briefly discuss our findings with respect to the use of AVR for joint exploration in general.

We based our initial design rationale mostly on related theory and works with modest involvement of field experts in an attempt to limit the influence of the SUD treatment context on the first iteration prototype. While we believe this approach indeed resulted in a general purpose implementation in the practical sense, it was only after more elaborately involving the SUD treatment context that we observed the importance of making a distinction between the practical and personal aspect of a joint experience, which is important to consider for any joint exploration application using AVR.

Our two main findings with respect to the SUD treatment use case, the role of embodied interaction and the shared physical space, are likely to apply to communication for joint exploration using AVR in general as well. For SUD treatment, these aspects were found to be a main priority within the communication, allowing us to determine their existence. For other contexts, these aspect might not be as important to the core of the interaction, but they can be considered to make the interaction more pleasant for example.

There are also the tools to consider outside of those necessary for communication on joint exploration. SUD treatment exercises were in that sense relatively easy to cover, as interpersonal communication is crucial for both. If a different application requires more elaborate ways to manipulate the VE, those tools need to fit next to the ones required for joint exploration. For example, the dashboard we presented in this research left limited room for buttons for actions outside of joint exploration.

Converting integral elements of the interaction to an embodied interaction, to make performing the interaction for joint exploration more similar to real life, can help free input space to fit application-specific tools. For example, the tablet dashboard could fit more tools for manipulating the VE outside of exploration as the room for more exploration-specific actions is no longer needed.

Furthermore, we argued that necessities for navigation could be generalized to include the utilization aspect of exploration as well, which does not seem to have caused conflicts within the SUD treatment use case. Therefore, it is possible that this generalization is applicable for joint exploration using AVR in general as well. Having clear guidelines for general AVR solutions for joint exploration benefits development of more specific solutions for other use cases, as development can then focus mainly on determining the additions of the specific use case to the general solution, as opposed to having to create a full solution from scratch.

6.6 Limitations

There are some limitations to this research and we are to be cautious against faulty interpretations or overgeneralization of our results. Arguably the most prominent limitation of our research is that it did not include one of the two main stakeholders: SUD clients with MBID. We excluded this stakeholder under the notion of their vulnerability in a yet to be refined application. We believe that this reasons remains valid to rely on treatment providers instead to provide insight on potential behaviour of clients.

We could have considered involvement of former clients that have completed their treatment, and as such be less vulnerable to SUD related cues presented in the exercises. Alternatively, we could have designed exercises with similar elements to joint exploration for SUD treatment, but without SUD-related elements, and included SUD clients with MBID in that regard. Both approaches should provide a more representable and complete insight for both the perspective of the client as well as the treatment provider, as results show that treatment providers are not always confident in their predictions on a client's behaviour or theirs with respect to a client's behaviour.

Moreover on stakeholders, treatment providers were identified as the other main stakeholder. They were identified as the people involved in the treatment process, including execution of related exercises. This group includes both therapists and sociotherapists. What was not considered was that within these two professions there are further specializations who can have separate requirements for such a system. This was made clear throughout both evaluations, where observing body posture of clients was deemed important by some participants in the first evaluation, but less so by participants in the second evaluation. Yet, participants in the second evaluation acknowledged that there were colleagues who might have deemed the aspect of observing body posture of the solution more important than they did.

Within the UCD iterations, we focused on developing and evaluating the interaction between users, and the appearance of elements within the VE itself was therein a second priority. This resulted in the use of a relatively simplistic, light, spacious, uncluttered, and unpopulated VE for both evaluations with the notable exception of the lunchroom with bar that was more decorated with items of varying sizes. In both evaluations the appearance of the VE, except for the bar, was indicated to be mostly not representable for envisioned exercises for SUD treatment. While it was not within the scope of the research to include fully representable VEs, it might be that our solution does perform differently in VEs that are more ecologically valid for SUD treatment, for example because small objects might be easily overlooked from the exocentric perspective or are hard to select in a pile of various other objects or the EPU avatar cannot easily manoeuvre in narrow spaces and has not the ability to sit down.

6.7 Future work

Aside from addressing the aforementioned limitations, we believe there are two main routes that can be taken to follow up this research: (i) further extend the research of AVR for SUD treatment or (ii) broaden the research to other contexts to further expand our knowledge on joint exploration using AVR in general.

For (i), the obvious follow up would be to further research and refine the interaction between client and treatment provider as IVRU and EPU respectively. A first step would be to address the aforementioned limited gestural cues of the EPU avatar and its inhuman appearance with regard to the identified lack of human qualities and expressing affect within the interaction of such an AVR setup.

Having a more refined interaction between two people, a logical next step would be to consider AVR interaction with groups consisting of more than two people, given the group exercises already provided in Tactus' treatment programs. Having more people to facilitate, certain ways of interaction might be rendered invalid and managing this group in both the virtual and physical space is a substantial challenge as well.

Furthermore, we can think about designing SUD treatment exercises specifically around the use of AVR as well, given the benefits IVR provides. For example, going beyond reality by incorporating a clone of the client or the fairy flower idea, as we established an interaction for which they can be a viable extension. This in contrast to executing existing exercises, that do not go beyond reality, such as role-play between treatment provider and client in a more ecologically valid environment. This requires looking into the placement of IVR technology within the treatment procedure as a whole as well.

For (ii), we already noted that different contexts have different needs and it is likely that there is no general AVR solution for joint exploration that can cover all contexts. Expanding the currently limited body of research on joint exploration using AVR would allow for better comparison of core aspects that individual contexts require an AVR solution to bring forward and how these should be accommodated in an attempt to ultimately solve joint exploration using AVR. Taking this research as a starting point, we see two different approaches: either continuing with the first iteration prototype or continuing with the second iteration prototype.

In the first UCD iteration we built a relatively general AVR solution. Evaluating this first iteration solution in various contexts other than SUD treatment and comparing the resulting requirements for a second UCD iteration allows for determining overlapping and unique requirements for each context, which can possibly be linked to traits of each context. Upon finding conflicting requirements, we know that a general solution is not possible. Alternatively one can start from the AVR solution of the second UCD iteration that is already more focused towards SUD treatment and evaluate that solution in various other contexts, allowing for the identification of overlaps and conflicts between contexts.

6.8 Conclusion

In this research we aimed to determine how to facilitate communication for joint exploration to support SUD treatment of people with MBID using AVR. For this aim we executed two user-centered design iterations, in which we built an initial prototype, based mostly on theoretical work with modest additional input from SUD treatment experts, and specialized that prototype further for SUD treatment of people with lower cognitive abilities with more elaborate input from SUD treatment providers.

Evaluating both iterations of the prototype with SUD treatment providers in exercises relating to the SUD treatment context provided us insight on how to facilitate the aforementioned communication. These evaluations allowed us to separate the necessities for enabling joint exploration using AVR and the implications the SUD treatment context put forward. However, key aspects of the interaction, such as embodied interaction, utilization of shared physical spaces, and presence requirements of both users, were found to be substantially shaped by the SUD treatment context.

This work seems to be the first to combine a focus on joint exploration using AVR and SUD treatment. Involved treatment providers showed enthusiasm for the possibilities of AVR for SUD treatment and its role in further developing treatment approaches.

References

[1] Ishani Aggarwal and Anita Williams Woolley. Do you see what i see? the effect of members' cognitive styles on team processes and errors in task execution. *Organizational Behavior and Human Decision Processes*, 122 (1):92–99, 2013.

[2] Wessel Ammerlaan. Group navigation in immersive virtual reality. 2020.

[3] Judith Amores, Xavier Benavides, and Pattie Maes. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1343–1348, 2015.

[4] Felipe Bacim, Eric D Ragan, Cheryl Stinson, Siroberto Scerbo, and Doug A Bowman. Collaborative navigation in virtual search and rescue. In *2012 IEEE Symposium on 3d user interfaces (3DUI)*, pages 187–188. IEEE, 2012.

[5] Albert Bandura. Self-efficacy mechanism in human agency. *American psychologist*, 37(2):122, 1982.

[6] Rüdiger Beimler, Gerd Bruder, and Frank Steinicke. Immersive guided tours for virtual tourism through 3d city models. In *Proc. the GI Workshop on Virtual Reality and Augmented Reality. p*, pages 69–75, 2013.

[7] Frank Biocca, Chad Harms, and Jenn Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pages 1–9, 2001.

[8] Frank Biocca, Chad Harms, and Judee K Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments*, 12(5):456–480, 2003.

[9] Cristina Botella, Rosa M Baños, Azucena García-Palacios, and Soledad Quero. Virtual reality and other realities. In *The Science of Cognitive Behavioral Therapy*, pages 551–590. Elsevier, 2017.

[10] 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):42, 2017.

[11] Doug A Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan Poupyrev. An introduction to 3-d user interface design. *Presence: Teleoperators & Virtual Environments*, 10(1):96–108, 2001.

[12] Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[13] Bill Buxton. Mediaspace–meaningspace–meetingspace. In *Media space 20+ years of mediated life*, pages 217–231. Springer, 2009.

[14] Marcio Cabral, Gabriel Roque, Douglas dos Santos, Luiz Paulucci, and Marcelo Zuffo. Point and go: Exploring 3d virtual environments. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 183–184. IEEE, 2012.

[15] Nicko R Caluya, Alexander Plopski, Christian Sandor, Yuichiro Fujimoto, Masayuki Kanbara, and Hirokazu Kato. Does overlay field of view in head-mounted displays affect spatial memorization? *Computers & Graphics*, 102:554–565, 2022.

[16] Emily Carl, Aliza T. Stein, Andrew Levihn-Coon, Jamie R. Pogue, Barbara Rothbaum, Paul Emmelkamp, Gordon J.G. Asmundson, Per Carlbring, and Mark B. Powers. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders*, 61:27 – 36, 2019. ISSN 0887-6185. doi: https://doi.org/10.1016/j.janxdis.2018.08.003. URL `http://www.sciencedirect.com/science/article/pii/S0887618518302469`. Virtual reality applications for the anxiety disorders.

[17] Brian L Carter and Stephen T Tiffany. Meta-analysis of cue-reactivity in addiction research. *Addiction*, 94(3):327–340, 1999.

[18] Elizabeth F Churchill and Dave Snowdon. Collaborative virtual environments: an introductory review of issues and systems. *virtual reality*, 3(1): 3–15, 1998.

[19] Carlos Coelho, JG Tichon, Trevor J Hine, GM Wallis, and Giuseppe Riva. Media presence and inner presence: the sense of presence in virtual reality technologies. In *From communication to presence: Cognition, emotions and culture towards the ultimate communicative experience*, pages 25–45. IOS Press, Amsterdam, 2006.

[20] Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. The cave: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–73, 1992.

[21] Ruth C Dalton, Christoph Hölscher, and Daniel R Montello. Wayfinding as a social activity. *Frontiers in psychology*, 10:142, 2019.

[22] Rudolph Darken and Barry Peterson. Spatial orientation, wayfinding, and representation. *Handbook of Virtual Environment Technology*, 2002, 11 2001. doi: 10.1201/b17360-24.

[23] Julia Diemer, Georg W Alpers, Henrik M Peperkorn, Youssef Shiban, and Andreas Mühlberger. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology*, 6:26, 2015.

[24] Paul Dourish. *Where the action is.* MIT press Cambridge, 2001.

[25] Jeff Dyck and Carl Gutwin. Awareness in collaborative 3d workspaces. Technical report, Technical Report, Computer Science Department, University of Saskatchewan, 2002.

[26] Marta Ferrer-Garcia, Olaya Garcia-Rodriguez, José Gutiérrez-Maldonado, Irene Pericot-Valverde, and Roberto Secades-Villa. Efficacy of virtual reality in triggering the craving to smoke: its relation to level of presence and nicotine dependence. *Stud. Health Technol. Inform*, 154:123–127, 2010.

[27] Marta Ferrer-Garcia, Olaya Garcia-Rodriguez, Irene Pericot-Valverde, Jin H Yoon, Roberto Secades-Villa, and Jose Gutierrez-Maldonado. Predictors of smoking craving during virtual reality exposure. *Presence: Teleoperators and Virtual Environments*, 21(4):423–434, 2012.

[28] Carlos Flavián, Sergio Ibáñez-Sánchez, and Carlos Orús. The impact of virtual, augmented and mixed reality technologies on the customer experience. *Journal of business research*, 100:547–560, 2019.

[29] Eelke Folmer, Isayas Berhe Adhanom, and Aniruddha Prithul. Teleportation in virtual reality; a mini-review. *Frontiers in Virtual Reality*, page 138.

[30] Mike Fraser, Steve Benford, Jon Hindmarsh, and Christian Heath. Supporting awareness and interaction through collaborative virtual interfaces. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 27–36, 1999.

[31] Borko Furht, editor. *Immersive Virtual Reality*, pages 345–346. Springer US, Boston, MA, 2008. ISBN 978-0-387-78414-4. doi: 10.1007/978-0-387-78414-4_85. URL https://doi.org/10.1007/978-0-387-78414-4_85.

[32] Lei Gao, Huidong Bai, Gun Lee, and Mark Billinghurst. An oriented point-cloud view for mr remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, pages 1–4. 2016.

[33] Martin Glinz and Roel J Wieringa. Guest editors' introduction: Stakeholders in requirements engineering. *IEEE software*, 24(2):18–20, 2007.

[34] Lauren Goode. Get ready to hear a lot more about 'xr'. https://www.wired.com/story/what-is-xr/. [Online; accessed 17-March-2022].

[35] Raphaël Grasset, Philip Lamb, and Mark Billinghurst. Evaluation of mixed-space collaboration. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pages 90–99. IEEE, 2005.

[36] Daniel Gromer, Max Reinke, Isabel Christner, and Paul Pauli. Causal interactive links between presence and fear in virtual reality height exposure. *Frontiers in psychology*, 10:141, 2019.

[37] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4021–4033, 2017.

[38] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. Facedisplay: towards asymmetric multi-user interaction for nomadic virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[39] Gengen He, Toru Ishikawa, and Makoto Takemiya. Collaborative navigation in an unfamiliar environment with people having different spatial aptitudes. *Spatial Cognition & Computation*, 15(4):285–307, 2015.

[40] Carrie Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271, 1992.

[41] Robin Horst, Sebastian Alberternst, Jan Sutter, Philipp Slusallek, Uwe Kloos, and Ralf Dörner. Avatar2avatar: Augmenting the mutual visual communication between co-located real and virtual environments. In *VISIGRAPP (2: HUCAPP)*, pages 89–96, 2019.

[42] Robin Horst, Fabio Klonowski, Linda Rau, and Ralf Dörner. The shared view paradigm in asymmetric virtual reality setups. *i-com*, 19(2):87–101, 2020.

[43] Weidong Huang and Leila Alem. Supporting hand gestures in mobile remote collaboration: a usability evaluation. In *Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction 25*, pages 211–216, 2011.

[44] igroup.org – project consortium. igroup presence questionnaire (IPQ) overview. `http://www.igroup.org/pq/ipq/index.php`. [Online; accessed 4-February-2021].

[45] Kisung Jeong, Jinmo Kim, Mingyu Kim, Jiwon Lee, and Chanhun Kim. Asymmetric interface: user interface of asymmetric virtual reality for new presence and experience. *Symmetry*, 12(1):53, 2020.

[46] Susanne Jul and George W Furnas. Navigation in electronic worlds: a chi 97 workshop. *ACM SIGCHI Bulletin*, 29(4):44–49, 1997.

[47] Shunichi Kasahara and Jun Rekimoto. Jackin head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM symposium on virtual reality software and technology*, pages 217–225, 2015.

[48] Roel Kerssemakers, Rob Van Meerten, H Vervaeke, and EA Noorlander. *Drugs en alcohol; Gebruik, misbruik en verslaving*. Bohn Stafleu van Loghum, 2015.

[49] Jan Kolkmeier, Emiel Harmsen, Sander Giesselink, Dennis Reidsma, Mariët Theune, and Dirk Heylen. With a little help from a holographic friend: The openimpress mixed reality telepresence toolkit for remote collaboration systems. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2018.

[50] Eike Langbehn, Paul Lubos, and Frank Steinicke. Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking. In *Proceedings of the Virtual Reality International Conference-Laval Virtual*, pages 1–9, 2018.

[51] Simon Langener, Joanne Van Der Nagel, Jeannette van Manen, Wiebren Markus, Boukje Dijkstra, De Fuentes-Merillas, Randy Klaassen, Janika Heitmann, Dirk Heylen, Arnt Schellekens, *et al.* Clinical relevance of immersive virtual reality in the assessment and treatment of addictive disorders: A systematic review and future perspective. *Journal of clinical medicine*, 10(16):3658, 2021.

[52] Joseph J LaViola Jr. A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32(1):47–56, 2000.

[53] Gun A Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pages 1–4. 2017.

[54] Jiwon Lee, Mingyu Kim, and Jinmo Kim. Rolevr: Multi-experience in immersive virtual reality between co-located hmd and non-hmd users. *Multimedia Tools and Applications*, 79(1):979–1005, 2020.

[55] Kevin Lynch. *The image of the city*, volume 11. MIT press, 1960.

[56] Nathan Matsuda, Brian Wheelwright, Joel Hegland, and Douglas Lanman. Vr social copresence with light field displays. *ACM Transactions on Graphics (TOG)*, 40(6):1–13, 2021.

[57] Dara Meldrum, Aine Glennon, Susan Herdman, Deirdre Murray, and Rory McConn-Walsh. Virtual reality rehabilitation of balance: assessment of the usability of the nintendo wii® fit plus. *Disability and rehabilitation: assistive technology*, 7(3):205–210, 2012.

[58] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12): 1321–1329, 1994.

[59] Daniel R Montello. *Navigation.* Cambridge University Press, 2005.

[60] D.R. Montello. *Spatial Cognition*, volume 7, pages 14771–14775. Pergamon, 12 2001. ISBN 9780080430768. doi: 10.1016/B0-08-043076-7/02492-X.

[61] Thi Thuong Huyen Nguyen, Cédric Fleury, and Thierry Duval. Collaborative exploration in a multi-scale shared virtual environment. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 181–182. IEEE, 2012.

[62] Max M North and Sarah M North. Virtual reality therapy. In *Computer-assisted and web-based innovations in psychology, special education, and health*, pages 141–156. Elsevier, 2016.

[63] Bashar Nuseibeh and Steve Easterbrook. Requirements engineering: a roadmap. In *Proceedings of the Conference on the Future of Software Engineering*, pages 35–46, 2000.

[64] Charles P O'Brien. Neuroplasticity in addictive disorders. *Dialogues in clinical neuroscience*, 11(3):350, 2009.

[65] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:114, 2018.

[66] Kaitlyn Michelle Ouverson, Carsten Scherber, Emily Oldham, and Stephen B Gilbert. What does "asymmetric vr" mean? a directed content analysis of co-located use of vr by users on reddit. *Frontiers in Virtual Reality*, page 154, 2021.

[67] Thomas D Parsons, Andrea Gaggioli, and Giuseppe Riva. Virtual reality for research in social neuroscience. *Brain sciences*, 7(4):42, 2017.

[68] Mark Peter, Robin Horst, and Ralf Dörner. Vr-guide: A specific user role for asymmetric virtual reality setups in distributed virtual reality applications. In *Proceedings of the 10th Workshop Virtual and Augmented Reality of the GI Group VR/AR*, 2018.

[69] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pages 1–5. 2017.

[70] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.

[71] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 6:5, 2019.

[72] Derek Reilly, Bonnie Mackay, Carolyn Watters, and Kori Inkpen. Planners, navigators, and pragmatists: collaborative wayfinding using a single mobile phone. *Personal and Ubiquitous Computing*, 13(4):321–329, 2009.

[73] 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.

[74] Giuseppe Riva. Virtual reality in clinical psychology. *Reference Module in Neuroscience and Biobehavioral Psychology*, 2020.

[75] Joan Sol Roo and Martin Hachet. Towards a hybrid space combining spatial augmented reality and virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 195–198. IEEE, 2017.

[76] Wendy A Schafer and Doug A Bowman. Evaluating the effects of frame of reference on spatial collaboration using desktop collaborative virtual environments. *Virtual Reality*, 7(3-4):164–174, 2004.

[77] Robert L Schalock, Sharon A Borthwick-Duffy, Valerie J Bradley, Wil HE Buntinx, David L Coulter, Ellis M Craig, Sharon C Gomez, Yves Lachapelle, Ruth Luckasson, Alya Reeve, *et al.* *Intellectual disability: Definition, classification, and systems of supports*. ERIC, 2010.

[78] Tomoyuki Segawa, Thomas Baudry, Alexis Bourla, Jean-Victor Blanc, Charles-Siegfried Peretti, Stephane Mouchabac, and Florian Ferreri. Virtual reality (vr) in assessment and treatment of addictive disorders: A systematic review. *Frontiers in neuroscience*, 13:1409, 2020.

[79] Helen Sharp, Anthony Finkelstein, and Galal Galal. Stakeholder identification in the requirements engineering process. In *Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99*, pages 387–391. Ieee, 1999.

[80] Jessica Simon, Anne-Marie Etienne, Stéphane Bouchard, and Etienne Quertemont. Alcohol craving in heavy and occasional alcohol drinkers after cue exposure in a virtual environment: The role of the sense of presence. *Frontiers in human neuroscience*, 14:124, 2020.

[81] Rigina Skeva, Lynsey Gregg, Caroline Jay, and Steve Pettifer. Views of practitioners and researchers on the use of virtual reality in treatments for substance use disorders. *Frontiers in Psychology*, 12:1852, 2021.

[82] 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.

[83] Aaron Stafford, Bruce H Thomas, and Wayne Piekarski. Comparison of techniques for mixed-space collaborative navigation. In *Proceedings of the Tenth Australasian Conference on User Interfaces-Volume 93*, pages 61–70. Citeseer, 2009.

[84] Mark Stefik, Daniel G Bobrow, Gregg Foster, Stan Lanning, and Deborah Tatar. Wysiwis revised: early experiences with multiuser interfaces. *ACM Transactions on Information Systems (TOIS)*, 5(2):147–167, 1987.

[85] Debra K Sullivan, Jeannine R Goetz, Cheryl A Gibson, Richard A Washburn, Bryan K Smith, Jaehoon Lee, Stephanie Gerald, Tennille Fincham, and Joseph E Donnelly. Improving weight maintenance using virtual reality (second life). *Journal of nutrition education and behavior*, 45(3): 264–268, 2013.

[86] Franco Tecchia, Leila Alem, and Weidong Huang. 3d helping hands: a gesture based mr system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*, pages 323–328, 2012.

[87] Wei-Peng Teo, Makii Muthalib, Sami Yamin, Ashlee M Hendy, Kelly Bramstedt, Eleftheria Kotsopoulos, Stephane Perrey, and Hasan Ayaz. Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation?–a narrative review of the literature. *Frontiers in human neuroscience*, 10:284, 2016.

[88] Santawat Thanyadit, Parinya Punpongsanon, and Ting-Chuen Pong. Observar: Visualization system for observing virtual reality users using augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 258–268. IEEE, 2019.

[89] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. Transceivr: Bridging asymmetrical communication between vr users and external collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 182–195, 2020.

[90] Mark H Trahan, Brandy R Maynard, Kenneth Scott Smith, Anne SJ Farina, and Yit Mui Khoo. Virtual reality exposure therapy on alcohol and nicotine: a systematic review. *Research on Social Work Practice*, 29 (8):876–891, 2019.

[91] JEL van der Nagel and Marion Kiewik. *Handleiding CGT+: Cognitief gedragstherapeutische behandeling van problematisch middelengebruik bij mensen met een lichte verstandelijke beperking*. Amersfoort: Stichting Resultaten Scoren, 2016.

[92] Neomi van Duijvenbode, Joanneke EL VanDerNagel, Robert Didden, Rutger CME Engels, Jan K Buitelaar, Marion Kiewik, and Cor AJ de Jong. Substance use disorders in individuals with mild to borderline intellectual disability: current status and future directions. *Research in Developmental Disabilities*, 38:319–328, 2015.

[93] JEL VanDerNagel. Minder drank of drugs handleiding: een behandeling voor patiënten met een lichte verstandelijke beperking en verslavingsproblemen. 2016.

[94] Craig Villamor, Dan Willis, and Luke Wroblewski. Touch gesture reference guide. `https://static.lukew.com/TouchGestureGuide.pdf`. [Online; accessed 10-March-2022].

[95] Jan Oliver Wallgrün, Mahda Bagher, Pejman Sajjadi, and Alexander Klippel. A comparison of visual attention guiding approaches for 360° image-based vr tours. pages 83–91, 03 2020. doi: 10.1109/VR46266.2020.1581092881445.

[96] Jia Wang, Rahul Budhiraja, Owen Leach, Rory Clifford, and Daiki Matsuda. Escape from meadwyn 4: A cross-platform environment for collaborative navigation tasks. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 179–180. IEEE, 2012.

[97] Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology*, 10:158, 2019.

[98] Tim Weissker, Alexander Kulik, and Bernd Froehlich. Multi-ray jumping: comprehensible group navigation for collocated users in immersive virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 136–144. IEEE, 2019.

[99] Karl Wiegers and Joy Beatty. *Software requirements*. Pearson Education, 2013.

[100] Jan M Wiener, Simon J Büchner, and Christoph Hölscher. Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation*, 9(2):152–165, 2009.

[101] Wikipedia contributors. Virtual reality sickness — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/wiki/Virtual_reality_sickness`, 2020. [Online; accessed 15-July-2020].

[102] Katie Witkiewitz and G Alan Marlatt. Relapse prevention for alcohol and drug problems: that was zen, this is tao. *American psychologist*, 59(4): 224, 2004.

[103] Katie Witkiewitz and G Alan Marlatt. Modeling the complexity of post-treatment drinking: It's a rocky road to relapse. *Clinical Psychology Review*, 27(6):724–738, 2007.

[104] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[105] Huahai Yang and Gary M Olson. Exploring collaborative navigation: the effect of perspectives on group performance. In *Proceedings of the 4th international conference on Collaborative virtual environments*, pages 135–142, 2002.

[106] André Zenner, Felix Kosmalla, Marco Speicher, Florian Daiber, and Antonio Krüger. A projection-based interface to involve semi-immersed users in substitutional realities. In *2018 IEEE 4th Workshop on Everyday Virtual Reality (WEVR)*, 2018.

[107] Xiaolong Zhang and George W Furnas. The effectiveness of multiscale collaboration in virtual environments. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pages 790–791, 2003.

A Design properties first iteration brainstorm ideas

The full design properties table of all 29 ideas is shown in Table 10 and Table 11.

Table 10: Design properties first iteration brainstorm ideas - part 1

	Initiation action			**Location of initiation**			**Freedom of movement EPU avatar**			**EPU avatar appearance**		**EPU avatar size**			
	single action	shape substance	create surface	chosen by IVRU	chosen by EPU	relative to object	move freely	move within limits	static location	humanoid	non-humanoid	human-sized	smaller than human	larger than human	depends on surface
Genie in bottle	X			X		X	X	X		X		X			
Magic paint		X	X	X			X		X	X		X			X
Magic remote	X			X		X	X			X		X			
Lamppost remote	X			X		X			X	X		X			
smoke bombs	X			X			X			X		X			
magic seeds	X			X				X	X	X			X		
magic clay		X		X			X		X	X			X		
magic goo	X			X			X			X		X	X		
oscar interaction	X			X		X			X	X		X			
shadow friend		X		X					X	X		X	X	X	
holodisk	X			X					X	X		X			
magic puddle			X	X					X	X					X
teletubby sun			X			X			X	X				X	
magic coin			X	X					X	X			X		
bubble friend		X		X		X	X			X		X	X		
origami friend		X		X			X		X	X			X		
pop-up book	X			X					X	X			X		
magic mirror			X	X					X	X					X
windmill screen			X			X			X	X				X	
create door	X	X		X			X			X		X			
magic scissors			X	X					X	X					X
magic flashlight			X	X					X	X		X	X	X	
butler bell	X			X		X	X			X		X			
charm wand	X						X			X		X			
magic hands	X		X	X			X		X	X		X	X		
therapist ball	X			X			X			X		X			
crystal ball			X	X					X	X			X		
beaty and the beast wand	X					X	X	X	X		X	X	X	X	
mr. bean entry	X				X		X			X		X			
clone yourself	X			X			X			X		X			

Table 11: Design properties first iteration brainstorm ideas - part 2

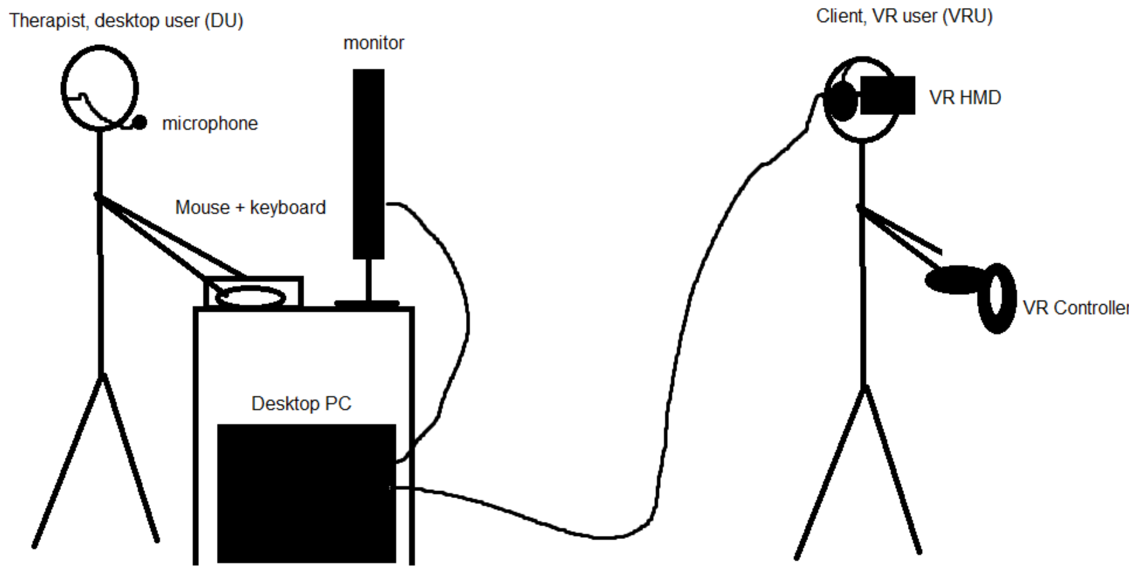
	requires carrying an object	leaves a trace	environmental dependent	Familiar interaction	Familiar EPU avatar	Initiation deniability	Limited number of uses
Genie in bottle	X			X			
Magic paint	X	X		X	X	X	X
Magic remote	X			X	X		
Lamppost remote	X		X	X	X		
smoke bombs	X			X	X		X
magic seeds	X	X		X			X
magic clay	X					X	
magic goo	X					X	
oscar interaction			X	X	X		
shadow friend	X						
holodisk	X				X	X	
magic puddle	X	X		X	X	X	
teletubby sun			X		X		
magic coin	X			X	X		
bubble friend	X			X		X	
origami friend	X			X		X	
pop-up book	X			X		X	
magic mirror	X			X	X	X	
windmill screen			X		X		
create door	X				X		
magic scissors	X	X			X		
magic flashlight	X			X	X	X	X
butler bell	X			X	X		
charm wand	X		X				
magic hands					X	X	
therapist ball	X				X		
crystal ball	X			X	X		
beaty and the beast wand	X		X				
mr. bean entry					X		
clone yourself							

B Interaction scenario first iteration prototype concept

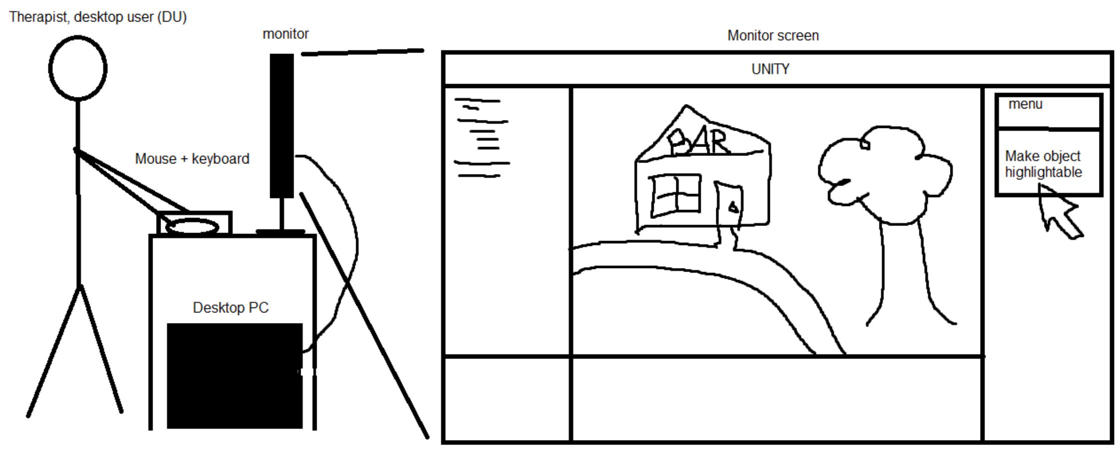
The following interaction scenario describes a possible interaction using the first iteration concept prototype. The scenario description is supported by various visuals that are referenced in line.

A rough description of one possible interaction with the system

Components: Desktop PC, Monitor, Mouse + keyboard, microphone, VR HMD, VR controller, sound-cancelling headset
Users: Therapist (Desktop user, DU), Client (VR user, VRU)
Goal scenario: practice with the 6Ds in a stimulating VR environment
Goal therapist: gain insight on the actions and reactions of the client during the exercise, help and give direction if anything is unclear for the client, intervene in the virtual environment if deemed necessary
Goal client: Practice with the 6Ds, (gain insights about themselves)
Scenario: VRU has to walk to the bus stop and take the bus home. Along the way to the bus stop, there are various stimuli and counter-stimuli



Technical preparations:
The therapist (or VR expert) prepares the virtual environment by giving all objects in the virtual environment, that they might want to (for example) highlight for the client to see, the property to make it highlightable in Unity (***1***). For simplicity they choose to make all objects in the virtual environment highlightable. They decide that they might change it to fewer objects in the future, but for the coming uses it is fine.



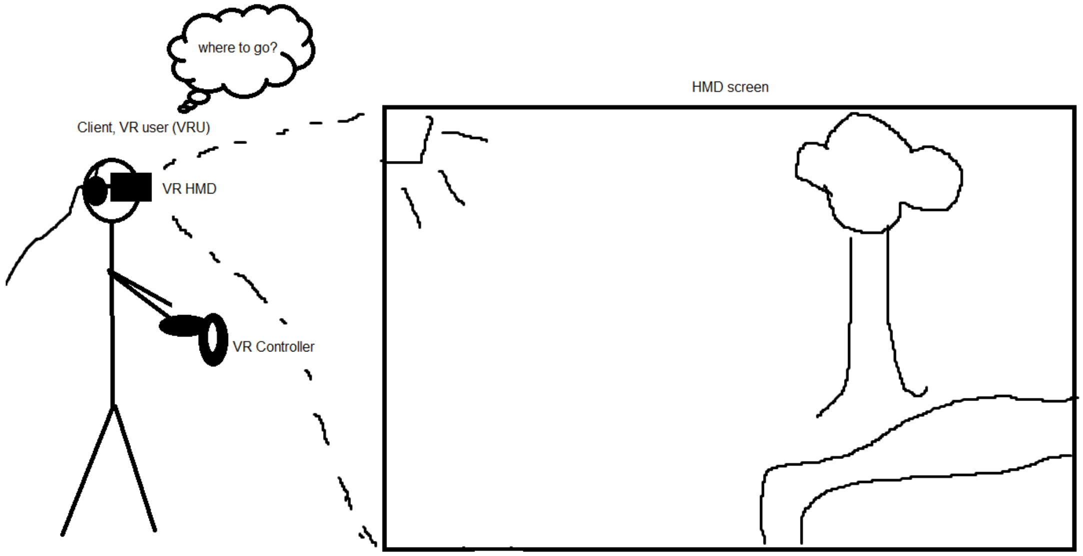
1

Before entering the virtual environment:
(This step assumes the client is introduced to the concept of VR and its components by the therapist)
Before entering the virtual environment, the client gets instructions on what to do in the virtual environment (go to bus stop, take bus home). The therapist starts the simulation before the client puts on the HMD to make it a screen rather than a blindfold.

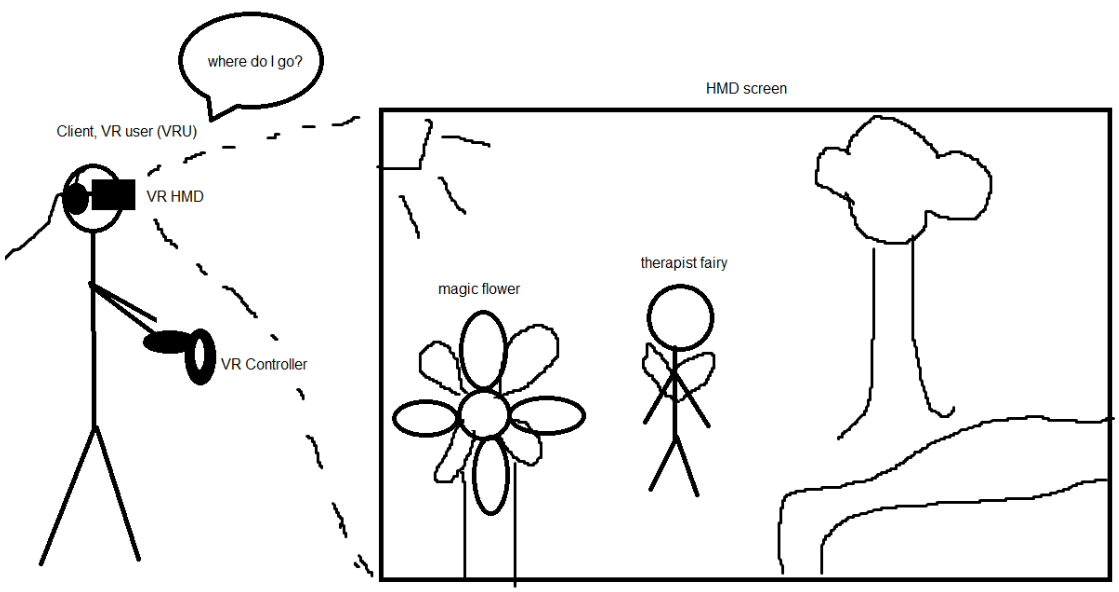
Entering the virtual environment:
The therapist helps the client put on the HMD. The client is still a bit overwhelmed by the technology, although it is not their first time in VR, and needs a moment to adjust.

In the virtual environment:

The client does not see a bus stop in its immediate vicinity and does not know where to go (***2***). Rather than wandering, they decide to ask the therapist and perform the action to get the therapist in the virtual environment with them (e.g. plant a magic seed, resulting in a magic flower from which a therapist fairy emerges) (***3***).

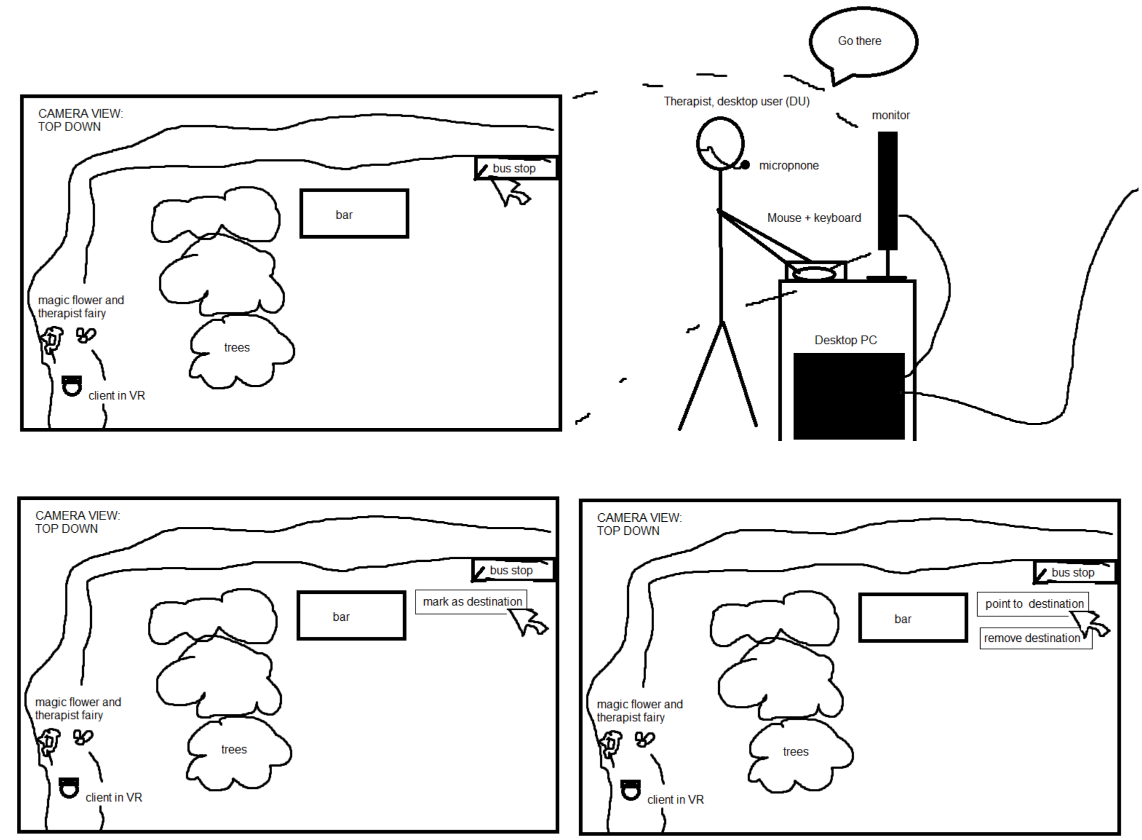


2

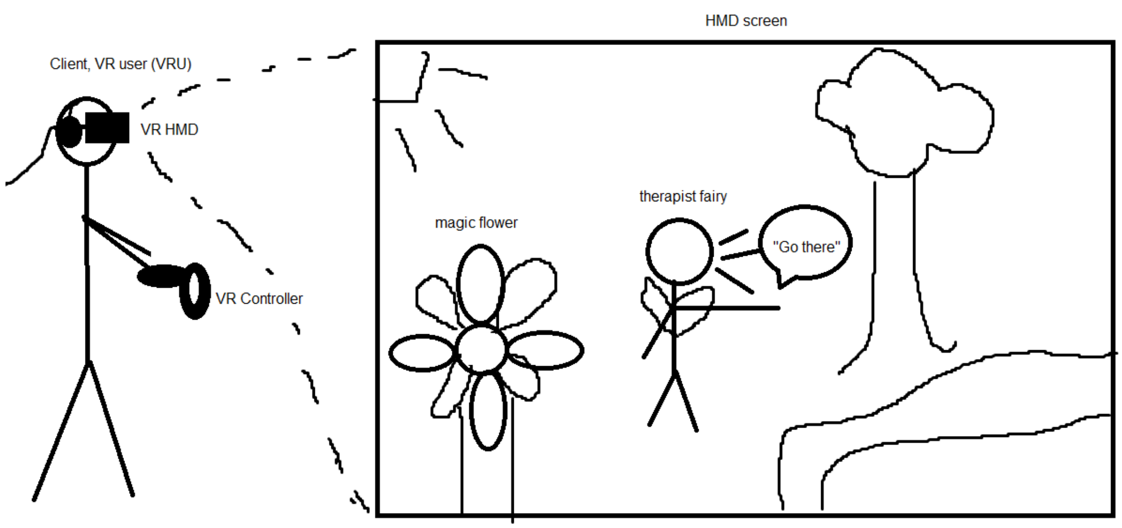


3

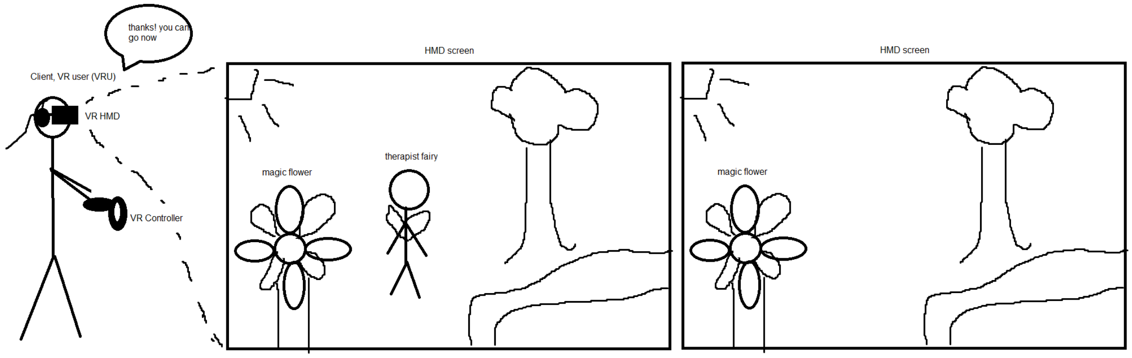
The client asks out loud where to go (***3***). The therapist gives a verbal reply via the microphone that is connected to the virtual agent (e.g. therapist fairy) for the client to hear in VR (***4***, top right). Additionally, the therapist selects the bus stop in their dashboard with an overview of the environment (***4***, top left) and marks it as the current destination (***4***, bottom left) and selects the option to let the agent point towards the destination (***4***, bottom right). The agent (fairy therapist) relays the answer and points towards the destination (***5***). The client now knows where to go and indicates that that was all. The embodiment disappears (and leaves a memento, a flower, in the example) (*6*).



4

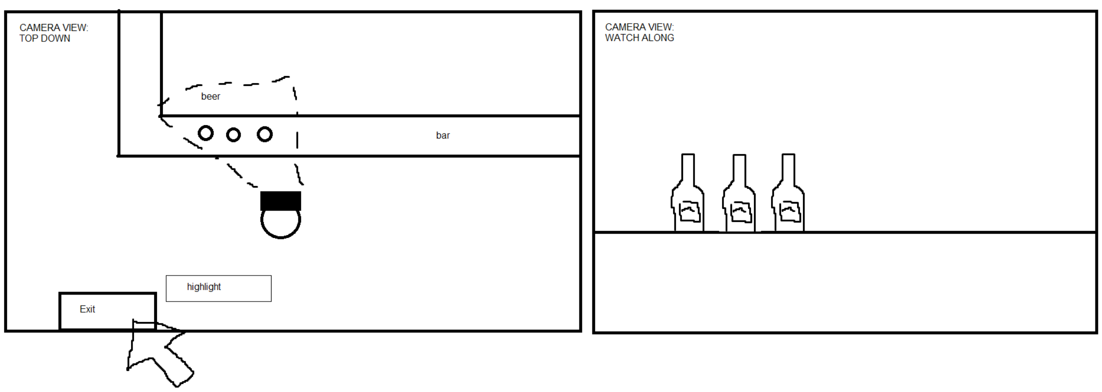


5



6

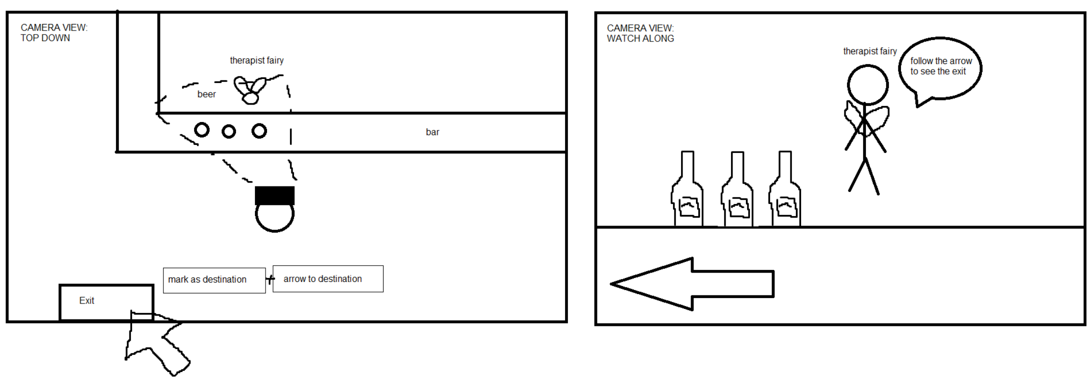
The client walks towards the given direction and encounters a bar. Not being able to resist urges, they enter the bar. The therapist sees this happening, takes a look inside the bar themselves, and tries to subtly suggest to leave the bar by highlighting the exit of the bar and checks whether the client sees it via the vision cone in the overview and double checks it by watching along with the client (*7*).



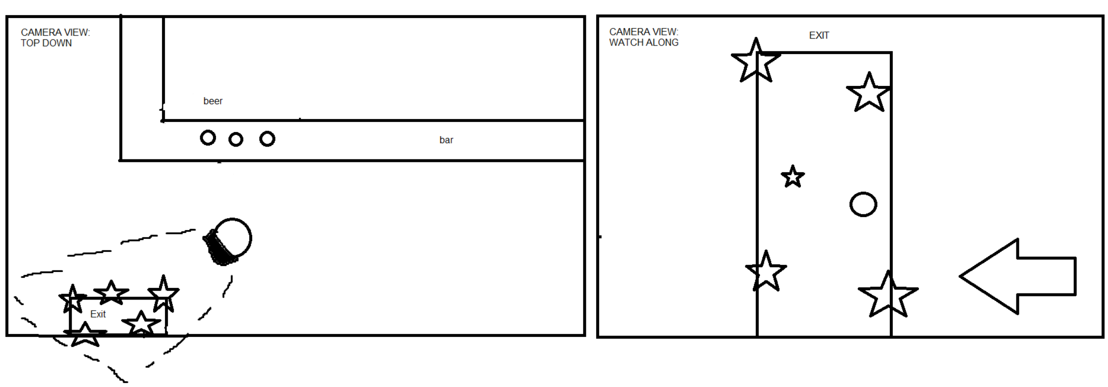
7

After determining that the client does not see the highlighted exit, the therapist decides to intervene by selecting a spot in the virtual environment and spawning in the embodiment (e.g. therapist fairy) and telling the client that they are going out of the bar (*8*). The therapist selects the bar exit as destination and selects to use a large arrow to point towards the exit (*9*). The client turns around and sees the sparkling exit and decides to leave the bar (*10*). Thereafter, the bus stop is quickly found and the scenario ends.

8



9

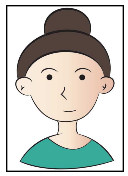


10

C Details first iteration evaluation

The next page contains the client persona as presented to participants of the first iteration evaluation. Table 12 provides an overview of the instructions of the research assistant for each encounter. Figure 39 provides a visual overview of the environmental elements that were present in the VE apart from the specific encounters. Lastly, Table 13 contains the original Dutch interview questions that participants were asked in the first iteration evaluation.

Cliënt-persona Thérèse



Achtergrondinformatie

Thérèse is een vrouw van 25 jaar. Ze is laagbegaafd en heeft een ernstige stoornis in het gebruik van alcohol. Ze heeft haar eigen appartement en woont daar alleen. Voorheen dronk ze alleen in groepsverband, maar sinds ze haar eigen appartement heeft en daar geen sociale controle ervaarde, begon ze ook thuis vaker te drinken met haar stoornis als gevolg. Ze heeft zelf besloten dat het zo echt niet meer langer kan en dat ze helemaal van alcohol af wil.

Ze drinkt vooral haar favorieten: Disaronno amandellikeur en Bailey's roomlikeur. Echter doet ze niet moeilijk over andere merken of soorten alcohol (wijn, bier, mixdrankjes) als haar favorieten niet beschikbaar zijn.

Voortgang behandeling

Thérèse is al een tijdje aanwezig in de kliniek en hier lijkt het al wat beter te gaan. Ze heeft geoefend in rollenspelen en is bekend met de 6A's van Minder Drank of Drugs. Deze VR-oefening is een voorbereiding op haar eerste verlof en jij bent haar coach. Ze heeft zelf al wat ervaring met de VR-technologie, maar heeft nog geen soortgelijke oefeningen gedaan.

Risicosituaties

Hoog risico

- Gezellig met vriendinnen thuis
- Op een terrasje of bij het uitgaan

Matig risico

- Verveling tijdens het wachten of als er niks te doen is
- Koffie (smaak en de geur), want "als het even kan gaat daar ook wat likeur doorheen"

Geen risico

- Bij haar ouders thuis, want "daar is sowieso geen alcohol aanwezig en ze zullen zich sowieso zorgen maken als ze mij met alcohol zien en dat wil ik niet"
- In de bibliotheek

6A's ingevuld

- ***Afstand***: Wegdraaien van de prikkel of naar buiten voor een andere geur
- ***Alternatieven***: Mocktails (*i.e.* alcoholvrije cocktails), gemberbier, koffiesiroop
- ***Afleiding***: Gamen op de computer
- ***Aangeven en bespreken***: Mijn moeder bellen
- ***Anders denken en anders doen***: -
- ***Applaus***: Mijn favoriete eten bestellen

Table 12: Research assistant instructions for each encounter

Encounter	**Instructions**
Between encounters	Follow instructions provided by the participant
Starting point	Ask the participant how to get to the bus stop Start following the instructions of the participant
Closed bar	Indicate the craving elicited by the sign Follow up instructions from the participant to handle this situation If the intermission is not too long, resume to follow prior navigational instructions. Otherwise ask where to go again if not already indicated by the participant
Accessible lunchroom with bar	Indicate that they would want to go take a look inside the lunchroom with bar No matter the instructions of the participant, walk into the lunchroom with bar Indicate that the space is more overwhelming than you thought and you are now uncertain how to handle the situation Follow up instructions from the participant to handle this situation Ask the participant where they should go again and start following the navigational instructions
Bus stop	Indicate that there are probably a few minutes left before the bus arrives and that you are not good at waiting around doing nothing and that you would normally have something to drink. Follow up instructions from the participant to handle this situation If the participant chooses for a 'create distance' tactic (again), deny as you are afraid to miss the bus that way.
Gas station (optional)	Despite any prior instructions of the participant, swiftly head on over to the gas station across the road. Instructions similar to the lunchroom with bar, but act slightly faster.

Figure 39: Neutral, tempting, and distracting elements as well as elements that suggest alternatives in the VE apart from a specific encounter.

Table 13: Original Dutch interview questions of the semi-structured interview for the first iteration evaluation

Related EQ	**Interview question**
EQ1-1	Wat is je eerste indruk van het systeem?
	Hoe vond je de oefening in het algemeen gaan?
	Was het systeem gemakkelijk in gebruik?
	Waren er in het programma mogelijkheden die je verraste?
	Waren er in het programma acties die beter waren dan de rest?
	Was er iets dat je duidelijk miste in het programma?
	Vond je dat het systeem was ingericht op gebruik door behandelaren?
	Wat zou er moeten veranderen voordat je het in de praktijk zou gebruiken?
	Voorzie je problemen bij het gebruik met echte cliënten?
	Hoe zou je het systeem anders gebruiken met echte clïenten?
EQ1-2 & EQ1-3	Hoe vond je de communicatie tussen jou en de VR-gebruiker gaan?
	Kon je gemakkelijk de intenties van de VR-gebruiker begrijpen?
	Had je het gevoel dat je intenties de VR-gebruiker gebereikte?
	(indien eerdere ervaring met AVR) Hoe was de communicatie anders dan je eerdere ervaring?
EQ1-4	Had je het gevoel dat je zelf ook voor een deel in de virtuele wereld aanwezig was?
	Of juist het gevoel dat de virtuele wereld vermenge met de echte wereld?
	Voelde de VR-gebruiker en het blauwe poppetje als dezelfde persoon?
	Voelde jij je verbonden met het oranje poppetje?
	Had je het gevoel dat je samen in de wereld zat?

D Digital third party resources

This appendix contains references to all digital third party resources that were used and did not get cited elsewhere within the thesis. The availability and existence of all resources were last checked on 09-March-2022, showing that one resource was no longer available and that resource is marked as such. The following tables provide the title, author, name of the published location, and url towards the resource where available. Table 14 provides an overview of Unity packages, Table 15 for animations, Table 16 for sounds, Table 17 for images, and Table 18 for icons. Providers of more than one resource are, in alphabetical order: Freesound[15], Mixamo[16], OpenMoji[17], Unity Asset Store[18], Unsplash[19], and Wikimedia Commons[20]. Providers of individual resources are mentioned within the tables themselves.

[15]Freesound: `https://www.freesound.org/`
[16]Mixamo: `https://www.mixamo.com/`
[17]OpenMoji: `https://openmoji.org`
[18]Unity Asset Store: `https://assetstore.unity.com/`
[19]Unsplash: `https://unsplash.com/`
[20]Wikimedia Commons: `https://commons.wikimedia.org/`

Table 14: Unity packages

Title	**Author**	**Obtained** from	**Source URL**
3D Character Dummy [v1.0]	Kevin Iglesias	Unity Asset Store	https://assetstore.unity.com/packages/3d/characters/humanoids/humans/3d-character-dummy-178395
Apartment Kit [v2.3]	Brick Project Studio	Unity Asset Store	https://assetstore.unity.com/packages/3d/props/apartment-kit-124055
Basic Motions FREE [v1.1]	Kevin Iglesias	Unity Asset Store	https://assetstore.unity.com/packages/3d/animations/basic-motions-free-154271
Cigarette Lighter PBR [v1.0]	devotid	Unity Asset Store	https://assetstore.unity.com/packages/3d/props/cigarette-lighter-pbr-106937
FREE Witchcraft and Wizardry Asset Pack [v1.0]	Ferocious Industries	Unity Asset Store	https://assetstore.unity.com/packages/3d/props/free-witchcraft-and-wizardry-asset-pack-141428
Polygon - FastFood [v1.0]	ANIMPIC STUDIO	Unity Asset Store	[NO LONGER AVAILABLE] https://assetstore.unity.com/packages/3d/props/polygon-fastfood-196151
Quick Outline [v1.1]	Chris Nolet	Unity Asset Store	https://assetstore.unity.com/packages/tools/particles-effects/quick-outline-115488
Same Material - Static Mesh Combiner [v1.1.4]	Lylek Games	Unity Asset Store	https://assetstore.unity.com/packages/tools/modeling/same-material-static-mesh-combiner-139565
SimplePoly - Town Pack [v1.0]	Gnome's Artworks	Unity Asset Store	https://assetstore.unity.com/packages/3d/environments/simplepoly-town-pack-62400
Tavern Bar Interior [v1.2]	3D Everything	Unity Asset Store	https://assetstore.unity.com/packages/3d/props/interior/tavern-bar-interior-46559
UCLA Wireframe Shader [v0.21]	The UCLA Game Lab	Unity Asset Store	https://assetstore.unity.com/packages/vfx/shaders/directx-11/ucla-wireframe-shader-21897

Table 15: Animations

Title	**Author**	**Obtained from**	**Source URL**
Standing clap from sitting	-	Mixamo	https://www.mixamo.com/#/?page=1&query=standing+clap+from+sitting
Waving	-	Mixamo	https://www.mixamo.com/#/?page=1&query=waving

Table 16: Sounds

Title	**Author**	**Obtained from**	**Source URL**
toaster oven or lift/elevator bell	azumarill	Freesound	https://freesound.org/people/azumarill/sounds/564623/
Sci-Fi sliding door (height adjustable chair sounds).wav	Pablobd	Freesound	https://freesound.org/people/Pablobd/sounds/525032/
Clap, Single, 7.wav	InspectorJ	Freesound	https://freesound.org/people/InspectorJ/sounds/404553/

Table 17: Images

Title	**Author**	**Obtained from**	**Source URL**
-	Mariah Hewines (@mariahhewines)	Unsplash	`https://unsplash.com/photos/DKD1K3HNq3g`
Bord open/gesloten - wit - kunststof	van Hattem Horeca B.V.	`https://www.vanhattemhoreca.nl`	`https://www.vanhattemhoreca.nl/klein-materiaal/tafelgerei/tafelnummers/bord-opengesloten-wit-kunststof-19718`
Bustijden_foto	-	`https://www.tilburg.com`	`http://tilburg.com/media/bustijden_foto.jpg`
Faux Neon Cocktails Sign	Will Murray (Willscrlt, `http://willmurray.name`)	Wikimedia Commons	`https://en.wikipedia.org/wiki/File:Faux_Neon_Cocktails_Sign.svg`
green and white drink with straw in clear drinking glass photo	LyfeFuel (@lyfefuel)	Unsplash	`https://unsplash.com/photos/_82CV9I-TP8`
long-stem wine glass photo	Marina Zaharkina (@minton)	Unsplash	`https://unsplash.com/photos/K8nr6rNDtUE`
Nederlands verkeersbord L3 bushalte	SVG version by Bouwe Brouwer	Wikimedia Commons	`https://commons.wikimedia.org/wiki/File:Nederlands_verkeersbord_L3_bushalte.svg`
Two mugs of brown liquid photo	kazuend (@kazuend)	Unsplash	`https://unsplash.com/photos/NmvMhov1sYc`
woman in gray tank top and blue denim jeans sitting on bed photo	Roselyn Tirado (@roselyntirado)	Unsplash	`https://unsplash.com/photos/cqAX2wlK-Yw`
yellow Volkswagen van on road photo	Dino Reichmuth	Unsplash	`https://unsplash.com/photos/A5rCN8626Ck`

Table 18: Icons

Title	**Author**	**Obtained from**	**Source URL**
Backhand index pointing right	Julian Grüneberg	OpenMoji	`https://openmoji.org/library/#emoji=1F449`
Clapping hands	Julian Grüneberg	OpenMoji	`https://openmoji.org/library/#emoji=1F44F`
Flag in hole	Vanessa Boutzikoudi	OpenMoji	`https://openmoji.org/library/#emoji=26F3`
Hollow red circle	Hilda Kalyoncu	OpenMoji	`https://openmoji.org/library/#emoji=2B55`
Man dancing	Johanna Wellnitz	OpenMoji	`https://openmoji.org/library/#emoji=1F57A`
Movie camera	Sina Schulz	OpenMoji	`https://openmoji.org/library/#emoji=1F3A5`
Notebook	Jonas Roßner	OpenMoji	`https://openmoji.org/library/#emoji=1F4D3`
Person walking	Johanna Wellnitz	OpenMoji	`https://openmoji.org/library/#emoji=1F6B6`
Wastebasket	Sina Schulz	OpenMoji	`https://openmoji.org/library/#emoji=1F5D1`
Waving hand	Julian Grüneberg	OpenMoji	`https://openmoji.org/library/#emoji=1F44B`
White square button	Ricarda Krejci	OpenMoji	`https://openmoji.org/library/#emoji=1F533`

E Coding scheme first iteration evaluation interview

Table 19 and Table 20 contain the coding structure that emerged from qualitative analysis of the interviews of evaluation 1 in table form. Note that the table is split into two separate tables for readability. Codes marked with a * were initialized a priori and codes without were established emergent.

Table 19: Codes per depth level, covering the top level code *System*. Codes marked with a * were initialized a priori and codes without were established emergent

Level 1	Level 2	Level 3	Level 4
System*	System in practice*	Usage with real clients*	Application as video game
		Required adjustments for use in practice	
		Placement within rehab process	
	System components*	User interface appearance*	
		Environment*	
		Virtual therapist*	
		Virtual therapist actions*	Applause*
			Pointing*
			Walk along*
			Walk to location*
			Wave*
		Main window*	Camera controls*
		Environment actions*	Attention grabber*
			Outline*
			Temporary landmark*
		Hover helpers*	
	System usage*	(Simultaneously) being a therapist	
		Break outside of VR for client	Turning off stimuli
		Focus on screen	Losing track of real client
		Habitation	
		Indicated needs	
		Indicated problems	
		Indicated superfluous	
		Naturalness interaction	
		Observing client body and face	
		Performing actions	
		Physical contact	
		Usage behind desktop	
		Usage patterns	
	Physical system setup	Physical distance	Physical interventions

Table 20: Codes per depth level, covering the top level codes *Joint collaboration*, *Virtual (co-)presence, Navigation, Real life exercises, and Experiment.* Codes marked with a * were initialized a priori and codes without were established emergent

Level 1	Level 2	Level 3	Level 4
Joint collaboration*	Awarenesscues*	Virtual client*	
		Client POV window*	
	Communication*	Communicating intentions	Communicating intentions therapist to client
			Understanding intentions client
		Verbal communication	
		Body language	
		Communicating feelings by client	
		Communicating praise	
		Conveying humor	
		Discussing new strategies	
		Easing the mood	
		Experienced distance in communication	
		Individual viewpoint	
		Knowing each other's whereabouts	Lost track of client location
			Lost track of virtual therapist location
		Understanding each other's viewpoint	
		Reflecting the client's emotions	
Virtual (co-)presence*	Virtual presence client	Being aware of therapist watching	
	Virtual presence desktop user		
	Virtual representation client via avatar		
	Virtual representation therapist via avatar	Experience virtual therapist as client	
	Taking on the VE together		
Navigation	Survey knowledge		
	Existing landmarks		
Real life exercises			
Experiment	Behaviour research assistant		
	Demo-part		
	Exercise-part		

F Details second iteration evaluation

Table 21 contain the original Dutch instructions given to the participants for each of the two exercises of the second iteration evaluation and their sub-parts. Figure 40 present the original Dutch questionnaire questions including the analog scale as how they were presented to the participants during the second iteration evaluation. Table 22 contains the original Dutch focus group questions and optional follow-up questions.

Table 21: Original Dutch instructions of each of the separate exercises

Exercise	Provided instruction
exercise 1 - part 1	Kijk samen wat rond en bespreek wat je ziet op dezelfde manier als dat je in het echt samen op stap zou gaan. Na ongeveer 5 minuten rondkijken krijgen jullie een nieuwe taak.
exercise 1 - part 2	Jullie moeten nu samen naar de bar die op jouw dashboard komt te staan navigeren. De bar is gemarkeerd door de rode pijl en marker.
exercise 2 - part 1	Jullie gaan zometeen de bar binnen en daar staan allemaal flesjes met verschillende dranksoorten. Alleen de cliënt kan deze flesjes oppakken en verplaatsen. Echter kun je kunt pas zien wat er in het flesje zit nadat deze is aangeraakt door de hand van de behandelaar. Probeer zoveel mogelijk flesjes op de bar te verzamelen. Er zijn er 11 totaal.
exercise 2 - part 2	We gaan nu een rollenspel spelen. De behandelaar zal zometeen de rol van barmedewerker aannemen die de cliënt graag alcohol wil verkopen. De cliënt die zichzelf speelt wil graag een drankje kopen en probeert met de situatie om te gaan. Nadat het rollenspel is uitgespeeld, bespreek hoe het rollenspel ging zonder de VR af te doen en probeer het eventueel opnieuw. Zorg er samen even voor dat de barmedewerker achter de bar komt te staan.

Figure 40: Original Dutch second iteration evaluation questionnaire questions and how they were presented to participants. Participants were asked to mark a point on the line for each question how they felt about the topic in regard to the completed exercises.

Table 22: Original Dutch second iteration evaluation semi-structured focus group questions

Dimension	Original Dutch focus group questions and optional follow-up questions (marked by a ∼)
Joint exploration (navigation) + collaboration in VEs	Hoe ging het samen navigeren door de stad? ∼In hoeverre konden jullie de intenties van de ander begrijpen? ∼Waren er elementen van het systeem die het samen navigeren in de weg zaten? ∼Waren er interactiemogelijkheden die jullie misten?
Joint exploration (utilization) + collaboration in VEs	Hoe ging het verzamelen van de flessen en het rollenspel? ∼In hoeverre konden jullie de intenties van de ander begrijpen? ∼Waren er elementen van het systeem die het samen verzamelen in de weg zaten? ∼Waren er interactiemogelijkheden die jullie misten?
Second iteration prototype	Hoe hebben jullie interactie via de tablet en positiesensoren ervaren? ∼Welke vorm van interactie had jullie voorkeur en waarom?
Presence and social presence	Hadden jullie het gevoel dat jullie samen de virtuele wereld aan het aangaan waren en waar kwam dat door? ∼Hadden jullie het gevoel samen in dezelfde virtuele wereld aanwezig te zijn en waar kwam dat door? ∼In hoeverre waren jullie bewust van de ander in de virtuele wereld? ∼In hoeverre hadden jullie het gevoel dat jullie de aandacht bij elkaar hadden? ∼Hadden jullie het gevoel meer bezig te zijn in de virtuele wereld of juist dat het jullie het gevoel hadden iets van buitenaf te bedienen en waar kwam dat door?
AVR for SUD treatment	Hoe zien jullie het gebruik van zo'n soort systeem voor oefeningen met echte cliënten voor jullie? ∼Waar voorzien jullie problemen bij interactie via het systeem voor therapieoefeningen? ∼Hoe bewust waren jullie van de echte omgeving?

G Coding scheme second iteration evaluation focus group

Table 23 contains the coding structure that emerged from qualitative analysis of the focus group results of the second iteration evaluation. Codes marked with a * were initialized *a priori* and codes without were established emergent.

Table 23: Coding scheme for focus group results. Codes marked with a * were initialized a priori and codes without were established emergent

Level 1	Level 2	Level 3
Joint exploration*	Exchanging strong wayfinding cues	Gestural communication EPU avatar
	Communicating intentions	
	Interaction possibilities EPU	
	Interaction for joint exploration	
Collaboration in VEs*	Awareness cues	General behaviour EPU avatar
		Virtual representation EPU via avatar
		Virtual representation IVRU via avatar
	Autonomy EPU	
	Non-avatar cues	
	Virtual collaboration	
AVR for SUD treatment*	Awareness real surroundings	
	Cognitive load	
	EPU as therapist	
	Observing VRU body language	
	Use with real clients	
Presence and social presence*	Social presence	social presence EPU as experienced by VRU
		presence EPU avatar
	Virtual presence EPU Virtual presence VRU	
Second iteration prototype*	Active tablet mode	Relative position IVR play area
	Passive tablet mode	Dashboard actions
	Habituation	
	Interpretation prototype 2	
Experiment specific	Interpretation experiment	
Miscellaneous	Possible design directions	