**The British
Psychological Society**

www.wileyonlinelibrary.com

Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape

Xueni Pan[1] and Antonia F. de C. Hamilton[2]* (iD)

[1]Department of Computing, Goldsmiths College, University of London, UK
[2]Institute of Cognitive Neuroscience, UCL, UK

As virtual reality (VR) technology and systems become more commercially available and accessible, more and more psychologists are starting to integrate VR as part of their methods. This approach offers major advantages in experimental control, reproducibility, and ecological validity, but also has limitations and hidden pitfalls which may distract the novice user. This study aimed to guide the psychologist into the novel world of VR, reviewing available instrumentation and mapping the landscape of possible systems. We use examples of state-of-the-art research to describe challenges which research is now solving, including embodiment, uncanny valley, simulation sickness, presence, ethics, and experimental design. Finally, we propose that the biggest challenge for the field would be to build a fully interactive virtual human who can pass a VR Turing test – and that this could only be achieved if psychologists, VR technologists, and AI researchers work together.

After many years of hype, virtual reality (VR) hardware and software is now widely accessible to consumers, researchers, and business. This technology offers the potential to transform research and practice in psychology, allowing us to understand human behaviour in detail and potentially to roll out training or therapies to everyone. The aim of this study was to provide a guide to the landscape of this new research field, enabling psychologists to explore it fully but also warning of the many pitfalls to this domain and giving glimpses of the peaks of achievement that are yet to be scaled. We consider both the advantages and limitations of VR technology, from a practical viewpoint and for the advance of theory.

In this study, we focus specifically on the use of VR for human social interactions, where a person interacts with another (real or virtual) person. VR is already widely used in studies of spatial cognition (Pine *et al.*, 2002) and motor control (Patton, Dawe, Scharver, Mussa-Ivaldi, & Kenyon, 2006) and these have been reviewed elsewhere (Bohil, Alicea, & Biocca, 2011). We also focus primarily on creating VR for the purpose of psychology experiments (rather than therapy or education; Rose, Brooks, & Rizzo, 2005). Note that we use the term VR to mean 'a computer-generated world' and not just 'things viewed in a head-mounted display', as the term is sometimes used. The latter includes things like 360°

*Correspondence should be addressed to Antonia F. de C. Hamilton, Institute for Cognitive Neuroscience, UCL, Alexandra House, 17 Queen Square, London WC1N 3AR, UK (email: a.hamilton@ucl.ac.uk).

video but excludes some augmented reality and non-immersive computer-generated systems which we cover here.

To frame the current study, we consider the world of VR as a new landscape in which the psychologist stands as an explorer, waiting at the edge of the map. We describe the challenges as mountains which this explorer will need to climb in using VR for research. First, we consider the foothills, describing the basic equipment which our explorer needs and mapping out the terrain ahead in a review of the practical challenges which must be considered in setting up a VR laboratory. Second, the Munros (peaks over 3,000 ft in Scotland) can be climbed by many with the correct equipment; similarly, we review the issues which may arise in implementing social VR scenarios and the best results achievable using current technologies. Finally, Olympus Mons (the highest mountain on Mars) has yet to be scaled; we consider the grandest challenge of creating fully interactive virtual people and make suggestions for how both computing and psychological theory must come together to achieve this goal. Throughout the paper, we attempt to give a realistic view of VR, highlighting what current systems can achieve and where they fall short.

Why bother?

Before even beginning on the foothills, it is worth asking why psychologists should use VR at all, and what benefits this type of interface might bring. As we will see, VR is not an answer to all the challenges which psychology faces, and there are many situations where VR is maybe a hindrance rather than a help. Nevertheless, VR has great promise in addressing some issues which psychology has recently struggled with, including experimental control, reproducibility, and ecological validity. These reasons help explain why many psychologists are now investing in VR and spending time and effort on making VR systems work.

A key reason to use VR in the study of social behaviour is to maximize experimental control of a complex social situation. In a VR scenario, it is possible to manipulate just one variable at a time with full control. For example, if you were interested in how race and gender interact to influence perspective taking or empathy, a live study would require four different actors of different races/genders – it is hard to assemble such a team, and even harder to match them for facial attractiveness, height, or other social features. With virtual characters, it is possible to create infinitely many combinations of social variables and test them against each other. This has proved valuable in the study of social perception (Todorov, Said, Engell, & Oosterhof, 2008) and social interaction (Hale & Hamilton, 2016; Sacheli *et al.*, 2015).

More generally, VR allows for good control of any interactive situation. For example, we might want to know how people respond to being mimicked by another person (Chartrand & Bargh, 1999), under social pressure (Asch, 1956) or to a social greeting (Pelphrey, Viola, & McCarthy, 2004). Social interactions are traditionally studied using trained actors as confederates who behave in a fixed fashion, and such approaches can be very effective. However, they are also hard to implement and even harder to reproduce in other contexts. Recently, there has been an increasing focus on reproducibility in psychology (Open Science Collaboration, 2015) and worries about claims that only certain researchers have the right 'flair' to replicate studies (Baumeister, 2016). Confederate studies in particular may be susceptible to such effects (Doyen, Klein, Pichon, & Cleeremans, 2012), or participants may be behave differently with confederates (Kuhlen & Brennan, 2013). All these factors make confederate interaction studies hard to replicate. In contrast, a VR scenario, once created, can be shared and implemented

repeatedly to allow testing of many more participants across different laboratories, which should allow for direct replication of studies as needed.

The traditional alternative to studying live social interactions is to reduce the stimuli and situation to simple cognitive trials with one stimulus and a small number of possible responses. For example, participants might be asked to judge emotion from pictures of faces (Ekman, Friesen, & Ellsworth, 1972) or to discriminate different directions of gaze (Mareschal, Calder, & Clifford, 2013). Such studies have provided valuable insights into the mechanisms of social perception, but still suffer from some problems. In particular, they have low ecological validity and it is not clear how performance relates to behaviour in real-world situations with more complex stimuli and a wider range of response options. Using VR gives a participant more freedom to respond to stimuli in an ecological fashion, measured implicitly with motion capture (*mocap*) data, and to experience an interactive and complex situation.

Finally, researchers may turn to VR to create situations that cannot safely and feasibly exist in the laboratory, including physical transformations or dangers which could not be implemented in real life. VR scenarios can induce fear (McCall, Hildebrandt, Bornemann, & Singer, 2015) and out-of-body experiences (Slater, Perez-Marcos, Ehrsson, & Sanchez-Vives, 2009). To give a social example, Silani *et al.* put participants in a VR scenario where they were escaping from a fire and had the opportunity to help another person, thus testing prosocial behaviour under pressure (Zanon, Novembre, Zangrando, Chittaro, & Silani, 2014). This type of interaction would be very hard (if not impossible) to implement in a live setting.

To summarize, VR can provide good experimental control with high ecological validity, while enabling reproducibility and novel experimental contexts. However, it is also important to bear in mind that the generalizability of VR to the real world has not been tested in detail. Just as we do not always know if laboratory studies apply in the real world, similarly we must be cautious about claiming that VR studies, where participants still know they are in an 'experimental psychology context', will generalize to real-world interactions without that context. The brief outline above demonstrates how VR systems have the potential to help psychologists overcome a number of research challenges and to answer important questions. However, there are also many issues which must be considered in setting up a VR laboratory and making use of VR in the study of human social behaviour. In the following sections, we review these challenges and consider if and how they can be overcome.

The foothills – how to use VR

Many researchers in psychology will have heard of VR, seen some demos, and tried on a headset. Fewer will have set up a VR laboratory or programmed a VR study. Here, we provide a short primer on the methods and terminology used in computing and VR. We focus particularly on how computing systems can take on the challenge of creating virtual characters (VCs) with behaviour that is contingent on the participant's actions. This requires information to flow both from the participant to the computer system and from the computer system to the participant (Figure 1). We consider the technology required for each in turn.

Hardware

Displaying the virtual world

There are a number of ways to visually display a computer-generated world to users, including head-mounted displays (HMDs), CAVE systems, augmented reality systems
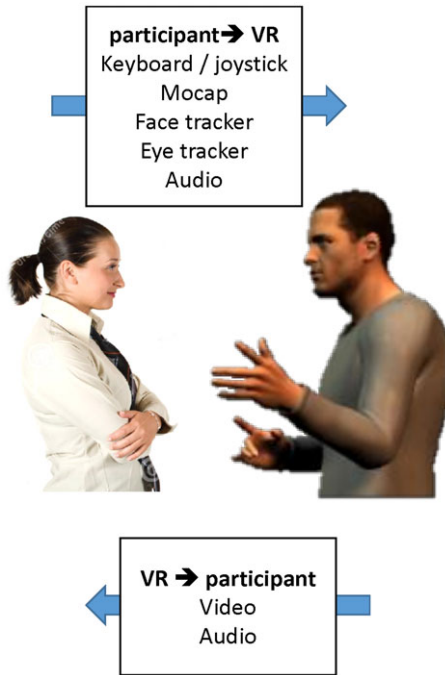
Figure 1. Participant interacting with a Virtual Human in virtual reality (VR). The VR system could take input from the participant through various channels and provide feedback mainly through video and audio.

which range from smartphones to headsets and finally projectors or desktop screens. Developments in this area are rapid and terminologies often overlap, but commonly used terms include immersive VR, mixed reality, and augmented reality. Immersive VR (discussed in more detail below) is typically experienced in an HMD which cuts the user off from the real world, while augmented reality places computer-generated items in the real world (sometimes allowing them to interact with the world) and mixed reality can include elements of both. Social interactions can be implemented across all these systems, and we review some of the most common approaches here.

The rise of VR in recent years has been mainly driven by the launch of several lightweight and affordable HMDs from many major consumer electronics companies (see Appendix S1). All these devices share one common feature: They provide an immersive experience (Slater, 2009) defined by (1) *3D stereo vision* via two screens – one in front of each eye; (2) *surround vision* – the real world is 'blocked' from your visual perception and as you turn your head you only see the 'virtual' world; and (3) *user dynamic control of viewpoint* which means that the user's head is tracked to update the display in real time according to where the user looks (Brooks, 1999). Implementing these three features together means that the visual information available in VR matches critical properties of the real world, where we have 3D vision all around and the visual scene updates with head movements. These immersive displays allow us to automatically respond to the computer-generated situation as if they were real [1] and are commonly described as *Immersive Virtual Reality* (IVR).

Despite the 'wow factor' of immersive VR displays, they do have some restrictions. First, the resolution of such devices is still relatively low compares to a standard computer

display, so it does not support studies requiring high-fidelity graphics (for instance, emotion reaction to subtle changes on the face). Secondly, as users are fully immersed with these VR displays, they are also 'cut-off' from the real environment. This makes it difficult to perform experiments involving interactions with real objects – although we could simulate the movement of the object in VR, the lack of generalized haptic devices means that it is difficult to completely replace real objects from our studies. Finally, in the context of social neuroscience research, immersive displays are not easy to combine with neuroimaging methods.

As an alternative to HMDs, CAVE VR systems can be very valuable. A VR CAVE typically has three or more walls with images projected onto each, giving a surrounding environment (Cruz-Neira, Sandin, & DeFanti, 1993). The user wears pair of shutter-glasses, which sync with the projector to generate 3D stereovision. Similar to HMDs, the glasses are tracked with six degrees of freedom (DoF) so the displays update in real time to render the perspective-correct view for the user. But unlike HMDs, in the CAVE the user can see through the glasses to view any real objects in the environment (including their own body). This could be a disadvantage for some applications as the glasses do not fully 'block' reality (i.e., one cannot fully embody someone else's body in the CAVE). However, it could also be an advantage for applications where the user can see and interact with real objects (e.g., a real driving wheel in a driving simulation) and can get real visual feedback of their own actions (e.g., hand actions in an imitation tasks). As it is challenging to implement virtual objects which respond to a user's actions (and almost impossible to create haptic feedback of objects in VR), it can be much simpler to allow the participant to access real objects in a CAVE.

There are also useful VR implementations which are even simpler than a CAVE. Some laboratories use VR content (i.e., animated 3D avatars) in a semi-immersive VR display, or even a non-immersive desktop display (Pfeiffer, Vogeley, & Schilbach, 2013; Sacheli *et al.*, 2015), sometimes coupled with 3D glasses. These could be considered as augmented reality rather than IVR and provide an interesting bridge between real and computer-generated worlds (de la Rosa *et al.*, 2015; Pan & Hamilton, 2015). Although it is arguable that immersion plays a significant role in triggering a realistic reaction in human participants, showing computer-generated virtual characters on a large screen can also be effective. In future, as the VR displays become higher fidelity and more wearable, and a wider variety of haptic devices become available, more and more experiments using virtual characters should be moving into the space of immersive VR. All these studies will also need to track the behaviour of the user, as we discuss in the following section.

Tracking the behaviour of the user – head, hands, body, face, and eyes
In addition to providing rich visual (and auditory) inputs to the participant, it is important for a VR system to be able to record and respond to the participant's behaviour. Many studies use the traditional methods of key-hits/mouse clicks to record a participant's behaviour, but richer measurement of behaviour can give excellent rewards. An appropriate system can allow researchers to record the motion of the hands, head, face, eyes, and body in varying resolution, which allows analysis of implicit and natural behaviours which may show much more subtle and interesting effects than traditional key-hit methods. Here, we review the different motion capture (*mocap*) systems available to VR researchers and the reasons for using them (see also Appendix S2).

First of all, most HMDs track head motion to update the visual display, and thus crude information on the direction of a participant's attention is available 'for free' in an HMD

system. Generally speaking, there are two types of HMDs: Some HMDs support both rotation and position (6DoF) tracking; others support rotation (3DoF) tracking only. An HMD with full 6DoF tracking provides a more immersive experience because a user can both walk in the space and turn their head. An HMD with only 3DoF only allows the user to turn their head but the body is fixed in space. In social interaction, we constantly adjust our position in relation to other people (for instance, we get closer to someone to share a private joke). With only 3DoF rotation tracking, this type of social signal cannot be supported.

Tracking the user's hand actions is often the next priority. Most high-end VR systems come with hand-held controllers that are tracked with 6DoF (e.g., Oculus Touch and the VIVE Controller) which means that if the user holds the controller, then his/her hands are tracked and can be represented in VR with 6DoF. However, such systems do not typically allow for variation in hand posture and gesture. In contrast, markless tracking systems (e.g., LeapMotion) and VR gloves (e.g., Manus VR) can permit natural conversational gesturing and richer hand motion. This also means that participants in an HMD can see a rendering of a hand moving with the correct timing, posture, and location to be their own hand, giving a stronger sense of embodiment in the VR world.

Full body tracking in VR can be achieved with a variety of systems, based on different combinations of cameras, magnetic markers, and inertia markers. It is commonly termed mocap (short for motion capture). We provide a more detailed review of these in Appendix S2, summarizing the types of technology available and the advantages/disadvantages of each. Important issues which need to be considered include the spatial and temporal resolution of a mocap system, whether the system is vulnerable to occlusion (optical systems) or to interference from other electronics (magnetic systems) and the latency with which the system can respond. All these issues are discussed in the Supporting Information. Different systems will be optimal for each of these functions so careful consideration should be given to the uses of body tracking data when setting up a VR laboratory.

Head orientation, sometimes in combination with head position, is often used to capture the gaze of the participant. Going further, it is possible to combine a VR system with eye-trackers to gain detailed gaze information, either with additional HMD compatible eye-tracking devices or using an HMD with build-in eye-trackers. Eye tracking in combination with responsive virtual characters on a screen has been used to yield interesting insights into the neural mechanisms of joint attention (Pfeiffer *et al.*, 2013). In addition, facial motion can be recorded with facial EMG or optical systems, but the latter are rarely compatible with an HMD because the HMD covers the upper half the face. This means that researchers interested in facial emotion might have to choose between recording the participant's facial motion with high fidelity and presenting stimuli in an HMD, but cannot easily do both with current systems.

As described above, technologies now exist to capture different aspects of participant' hand, face, and body actions in the context of VR research. However, capturing all aspects of behaviour at once remains a challenging problem, and these technical limitations impose critical constraints on what psychology studies can be done. Despite the many challenges in the domain of mocap, there are many reasons why we believe that rich capture of the human behaviour is valuable for social interaction research.

First, mocap data can be used to generate realistic yet well-controlled virtual character animations stimuli (de la Rosa, Ferstl, & Bülthoff, 2016). For instance, de la Rosa and colleagues used mocap data to create a range of stimuli showing the actions which the actors actually performed ('fist bump' and 'punch'), but also ambiguous stimuli blending

the two animations. Perception of these blended mocap stimuli could be tested in an adaptation context (de la Rosa *et al.*, 2016) or in the context of different facial identities (Ferstl, Bülthoff, & de la Rosa, 2017). Using a similar method, (Sacheli *et al.*, 2015) applied the same animation clip to virtual characters with different skin colours (white and black) and found that a stronger interference effect on participant's motion from an in-group VC as compared to the outgroup one. These studies illustrate the value of using mocap and virtual character technology to create experimental stimuli with precise control.

Second, capturing participant motion means the VR environment can be programmed to be responsive in real time, with both embodiment (see below) and realistic interactions between the participant and other objects or characters. For instance, knowing the participant's head location means that a virtual character can be programmed to orient their head and/or gaze towards participant's head (Forbes, Pan, & Hamilton, 2016; Pan & Hamilton, 2015) and to maintain an appropriate social distance by stepping back or forward (Pan, Gillies, Barker, Clark, & Slater, 2012). The ability to link the behaviour of a virtual character to the participant in real time also facilitated a series of studies on mimicry in VR, where the virtual character copies participants' head movements (Bailenson & Yee, 2007; Verberne, Ham, Ponnada, & Midden, 2013), or both head and torso movements (Hale & Hamilton, 2016).

Finally, motion capture allows the researcher to record natural and unconstrained behaviours. This permits measures such as proxemics (McCall & Singer, 2015), approach as a measure of trust (Hale, Payne, Taylor, Paoletti, & Hamilton, 2017), and imitation (Forbes *et al.*, 2016; Pan & Hamilton, 2015). For instance, McCall & Singer conducted a study where participants in an HMD explored a virtual art gallery while two other VCs (representing people the participant believed to be fair or unfair) remained in fixed locations. The position and orientation of the participant's head provided an implicit 'proxemics' measure of how much they liked each VC (McCall & Singer, 2015). Similarly, participants can be placed in a virtual maze where they can approach different VCs for advice to find a way out. The choice of which virtual character for advice and whether they followed the advice (i.e., which door they then chose to go through) provided an implicit measure of trust (Hale *et al.*, 2017). Other studies recorded participants' hand position during their interaction with a virtual character as a measurement of imitation and found that typical adults automatically imitated the virtual characters, but participants with autism spectrum conditions imitated less (Forbes *et al.*, 2016; Pan & Hamilton, 2015). These studies illustrate the use of VR to record implicit social behaviours which may be more revealing than traditional key-hit measures.

To summarize, the hardware required for VR laboratory typically comprises both visual displays and motion capture systems, with a wide variety of solutions available for different tasks and contexts (See also Supporting Information). While there are a number of complex choices involved in getting the right hardware, using VR together with advanced mocap solutions yields substantial benefits in capturing valid data and creating realistic VR. However, the hardware must always be combined with appropriate software to create a psychological experiment, and so we turn next to the domain of software.

Software

The software package which implements a VR experience with virtual characters is the core component which creates a social experience and an immersive world. A variety of commercial and open-source packages are available, but creating an immersive virtual interaction within these can be a major undertaking. The present section does not attempt

to review all current software packages as this is a rapidly changing field (see Appendix S3), but rather we provide an introduction to the terminology of the field and the critical issues which must be considered in developing a VR scenario in any software package.

To introduce the reader to the range of what VR software can (and cannot) do, we imagine a landscape of possible VR systems (Figure 2). In this landscape, we distinguish between different types of virtual interaction on two dimensions – the level of interactivity between participant and computer system (y-axis), and the level of graphical realism which participant's experience (x-axis). At one extreme, we can consider the movie *Avatar* (Cameron, 2009) which has photorealistic characters but there is no potential to interact with them as the story progresses (Figure 2A). At the other extreme, the computer game PacMan has very simple pixelated characters which are highly responsive to both key-hits and to each other (Figure 2B). This illustrates the difference between graphical realism and interactivity.

PacMan also provides a good example of an important distinction in the domain of virtual characters (VCs). Pacman itself is an *Avatar*, a character who is fully controlled by a human being. The ghosts who chase pacman are described as *Agents* – characters who are fully controlled by algorithms (in Computer Games, this is often called 'Non-Player Character', or NPC). While the word 'avatar' is sometimes used for any character that looks 'computer generated', or any graphical representation of the user in the virtual world, we argue that it should technically be reserved only for those characters which are fully controlled in real time by another person. In between the two extremes of Avatar and Agent lies the interesting domain of quasi-agents – characters which are partly autonomous and partly controlled by a human. These are increasingly widely used in
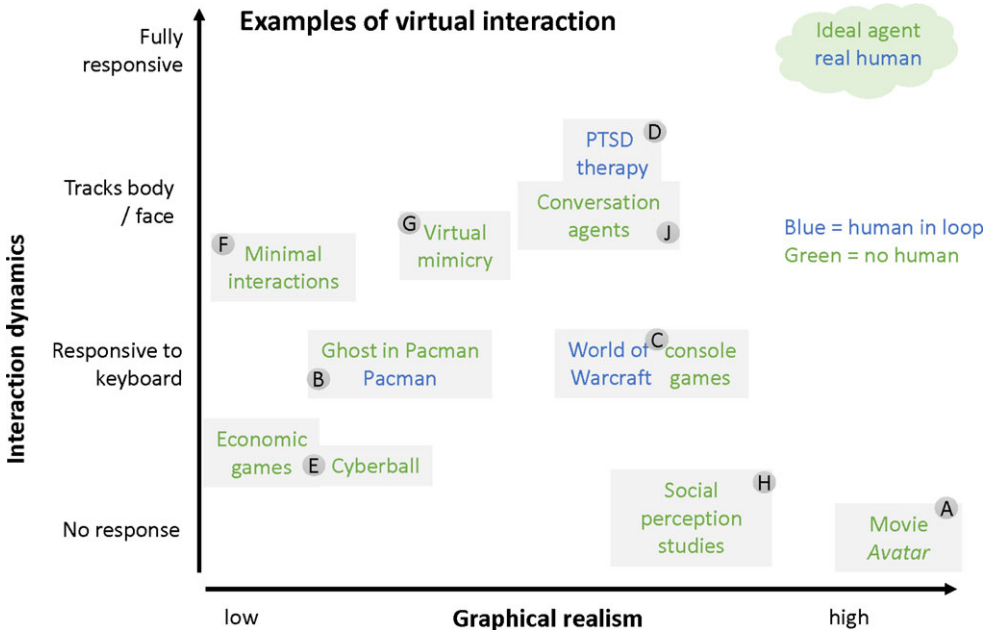
Figure 2. The landscape of virtual interaction. We distinguish current technologies on two axis – graphical realism and interaction dynamics. Examples to match each letter are given in the text.

therapy contexts and gaming contexts. For example, characters in most popular computer games and online worlds (e.g., FIFA game, World of Warcraft) will show some behaviours automatically but other behaviours only when the user hits a key (Figure 2C). In therapies which use VCs (Pan *et al.*, 2012; Rizzo *et al.*, 2015), a conversation agent is typically used in which some actions (e.g., gaze, proxemics, gestures, smiles) are pre-programmed while other aspects of the conversation are controlled by a therapist who listens and watches, then pressing keys on a keyboard to trigger specific events (Figure 2D). Such systems are described as 'Wizard of Oz' systems because the behaviour appears to come from the virtual character but is actually driven by a human 'wizard'.

Common psychological studies can also take their place in the interaction landscape. Virtual games such as Cyberball (Williams & Jarvis, 2006) or multi-round economic games (Hampton, Bossaerts, & O'Doherty, 2008; Yoshida, Seymour, Friston, & Dolan, 2010) in which participants play against an algorithm fall on the lower left of the plot (Figure 2E). Such games are interactive at a fixed time frame (there are only some time points where a participant can press a key) and which have minimal graphics, but which nevertheless can be very valuable in psychological research. Economic games can also be built into more elaborate interfaces with virtual characters to determine how non-verbal behaviours change decision-making (Gratch, Nazari, & Johnson, 2016).

Fully responsive systems have been built using very minimal interfaces, such as a single mouse and a pair of boxes moving on the screen (Auvray, Lenay, & Stewart, 2009), and these can allow the study of minimal interactions (Figure 2F). Richer responsive systems that implement mimicry are rare but can be used (Figure 2G). In contrast, studies of social perception commonly use computer-generated characters (e.g., Jack & Schyns, 2015; Todorov *et al.*, 2008) but do not allow participants to have any interaction with the figures they see (Figure 2H). Thus, mainstream psychological studies tend to remain close to the *x*-axis or close to the *y*-axis in the virtual interaction landscape. We argue below that there is both potential and more value in moving further and creating psychological studies with greater responsiveness and realism.

The state-of-the-art and current limits
In our interaction landscape, we place live humans in the top-right hand corner with perfect graphical realism and full responsiveness. An ideal VC system would inhabit the same space, giving the experimenter both real-time responsiveness and high realism. Here, we review how far current systems have got towards this goal.

Autonomous or semi-autonomous virtual agents are computer-generated characters which can engage in realistic interactions with a participant. These systems, built up from decades of work, represent the cutting edge of creating realistic virtual agents. They can register the gestures, body motion, and gaze of a user and generate in real time both the verbal and non-verbal cues required to effectively communicate with the user, giving a startling impression of realism. Some semi-autonomous agents have been designed to enable therapy for conditions such as phobias (Pan *et al.*, 2012) and PTSD (Rizzo *et al.*, 2015; Figure 2D). Such systems typically function with a human therapist acting as the Wizard of Oz, both to monitor the progress of the therapy and to select appropriate behaviours for the VC to show. Similar systems have been built to explore processes of negotiation (Gratch, Devault, & Lucas, 2016) or mimicry (Hasler, Hirschberger, Shani-Sherman, & Friedman, 2014).

Fully autonomous agents can also be built, in which a VC can conduct a brief conversation with a user with no human control (Figure 2J). Such systems typically use an

array of sensors to determine what the user is doing, including motion capture (see above), speech recognition, face capture (Baltrusaitis, Robinson, & Morency, 2016), acoustic speech analysis (Eyben, Wöllmer, & Schuller, 2010). Inputs from these systems are fed into an AI model which determines the user's goals and provides appropriate outputs. The outputs must be translated into speech and gestures using tools such as BEAT (Cassell, Vilhjálmsson, & Bickmore, 2001) or the Virtual Human Toolkit (Gratch & Hartholt, 2013). Each system tends to work within a limited domain of knowledge, as defined by the semantic model, but can be fairly effective within this domain.

We review here some examples of the state-of-the-art in this area. First, the SEMAINE project created four Sensitive Artificial Listeners (SALs) with different personalities: the aggressive Spike, the cheerful Poppy, the gloomy Obadiah, and the pragmatic Prudence (Schröder, 2012). All were autonomous agents and can interact with users in real time, without a human operator. Second, the USC Institute for Creative Technologies developed the SimSensei Kiosk, Ellie, an autonomous virtual human interviewer able to engage users for a 15–25 min interaction where they would feel comfortable to share personal information (DeVault *et al.*, 2014). Ellie is designed to automatically assess user's mental health status and identify issues such as depression, anxiety, or post-traumatic stress disorder. More recently, The ArticuLab at Carnegie Mellon University has designed social aware robot assistant (SARA), a virtual character who is able to recognize both non-verbal (visual and vocal) and verbal signals and utilizes AI to form her answer (Zhao, Sinha, Black, & Cassell, 2016). SARA's AI is motived by two goals: task (answering questions, such as help the user to find directions) and social (maintaining a positive and engaging relationship with the user). After a response is formed with its AI model, both verbal and non-verbal behaviours are generated to allow a realistic interaction with the user. All of these systems are focused on generating an emotional connection with the user.

A slightly different approach is taken by the team who developed an artificial agent named Billie, which can converse with users using both gesture and speech. Billie's behaviour is driven by systems based on cognitive models of motor control and mentalizing, including principles of active inference (Kahl & Kopp, 2017). The system is able to create common ground in a simple communication task about personal organization and diary entries and will ask for clarification if it does not understand or interrupt politely if the user goes off-topic. As the implementation draws on ideas from cognitive psychology, it also provides an example of how virtual agents can be used to test psychological theories (Kopp & Bergmann, 2017).

The systems described above are at the cutting edge of current virtual agents, but still have some limitations. Each system typically remains tied to a very specific social contexts and can typically discuss only one or two pre-trained topics. The behaviours and gestures which can be recognized and produced must be carefully specified by the researchers, and most systems use only a small subset of the behaviours of a real person. Learning and adaptability is not yet built in. Finally, most of these fully autonomous agents were only tested with a simple non-immersive VR display (i.e., a computer screen) so that participants' facial expression can be tracked. To test the full effect of those autonomous agents and really compare them to real-world social interaction, new ways of integrating the tracking technology enabled multimodal approach and the immersive display need to be explored. Overall, creating a general and fully responsive virtual agent remains a very large challenge for the future (see below).

The Munros – challenges in the implementation of VR

In the section above, we have set out the basic requirements of a social VR system and considered where current psychological research fits in relation to such a system. Our map of the VR landscape allows us to navigate the space of possible studies, but there are still mountains to climb and potential pitfalls to avoid. Here, we set out to guide the novice VR researcher beyond these basic foothills and provide the information needed to scale the Munros. That is, we describe the technical and practical challenges which face a researcher setting up a new VR laboratory, together with some guidance for solving them. As before, we focus on circumstances where research into human social interaction is likely to be affected.

The challenge of self-embodiment

When a participant puts on an HMD, they lose sight of their own body. In some experimental contexts, this does not really matter – a participant in an MRI scanner also has very limited visual input from their own body and can still perform many psychologically useful tasks. However, there is evidence that lack of embodiment can lead to worse performance on a variety of tasks which make use of the self-image such as mental rotation (Steed, Pan, Zisch, & Steptoe, 2016). Giving a participant a realistic and believable experience of having a body can be critical to many studies. This can be achieved through visual-proprioception synchrony (i.e., the virtual body or body parts are where you expect your body to be), visual-motor synchrony (as you move your body, the virtual body moves the same way), or visual-tactile synchrony (as you experience touch on part of your body, you see the same virtual body parts being touched at the same time).

Visual-tactile synchrony has been widely used without VR, in the rubber hand illusion (Botvinick & Cohen, 1998), and the enfacement illusion (Tsakiris, 2008). The same principles can be applied in VR, using an HMD and a live-feed video from a mannequin being synchronously stroked as the participant's own body (Petkova & Ehrsson, 2008), or a virtual arm being synchronously stroked as the participant's own arm (Slater, Perez-Marcos, Ehrsson, & Sanchez-Vives, 2008). More recent studies use visual-motor synchrony (usually in combination with visual-proprioception and sometimes visual-tactile synchrony) to create feeling of embodiment. This typically means that the participant's movements must be captured with a motion tracking system and displayed in the VR world in real time and in the appropriate spatial location. Wearing an HMD, participants could look down to see their own virtual body and observe their virtual body moving in time with their real body (Slater, Spanlang, Sanchez-Vives, & Blanke, 2010). To enhance the illusion, often a mirror is used so participants see 'themselves' moving in the mirror in VR.

Once embodiment is established, it is possible to manipulate the participant's sense of body in various ways. This includes changing the spatial location of the body (Slater *et al.*, 2009), the age of the body (Banakou, Groten, & Slater, 2013), and the race of the body (Peck, Seinfeld, Aglioti, & Slater, 2013). These methods open up a rich vein of research for psychologists to investigate the sense of self and we recommend (Maister, Slater, Sanchez-Vives, & Tsakiris, 2014) as a review of this area. In practical terms, a variety of software solutions are available to implement embodiment, but their success depends critically on the quality of the motion capture and the time lags in the computers. For a full discussion about the technical setup of VR embodiment, we recommend (Spanlang *et al.*, 2014).

The challenge of the uncanny valley

The concept of the uncanny valley was introduced by Mori, MacDorman, and Kageki (2012) who suggested that there is a non-linear relationship between how human-like a robot or virtual character looks, and how people perceive it. Specifically, he proposed that characters which look nearly-but-not-quite human are judged as uncanny and are aversive. More systematic studies suggest that an uncanny valley exists for still images morphed between a human and robot appearance (MacDorman, 2006) but is not always present when characters are animated (Piwek, McKay, & Pollick, 2014). It may be that uncanniness arises when there is a disparity between the appearance of a character and the way in which it moves (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012), such that a highly photorealistic human-like figure moving in a jerky fashion would be perceived as more uncanny than a cartoon-like figure moving in the same way. These studies suggest that a key requirement for creating believable virtual characters is to use smooth, realistic motion and that it is not essential to use highly photorealistic virtual characters. A more detailed review of this issue is provided here (de Borst & de Gelder, 2015).

The challenge of simulation sickness

Many users experience nausea during their VR experience, especially with HMD VR systems. However, not all users experience simulation sickness to the same extent, and certain applications cause more severe nausea than others. The main contributor of simulation sickness is the conflict between the visual and vestibular systems – where the user perceives they are moving with their eyes but not their body – which is the opposite of the motion sickness felt on a car or train. One simple fix to this is to use 'physical navigation' in which the user can move around a large space on the same physical scale as the VR world, keeping user's visual and body motion consistent. This is often referred to as 'room-scale VR', and its use in research is constrained primarily by the size of the room available to the researchers. Other contributing factors to simulation sickness in VR HMDs including eye strain (the displays are very close to your eyes), latency (as you turn your head, the image has a delay in updating), and high contrast images. The impact of these can be reduced by changing the design of the VR environment, for example, limiting motion speed or reducing the intensity of optic flow as the user moves.

Because many factors from both hardware and software contribute to simulation sickness, it is hard to estimate the percentage of participants affected. In a recent study where participants moved through a virtual maze with an HMD device (Hale *et al.*, 2017; study 2), three of 24 participants, or 12.5% terminated the task before completion due to simulation sickness. A large-scale study recruited 1,102 participants to go through an HMD experience and found that the dropout rate was 6.3% for 15 min and 45.8% for 60 min (Stanney, Hale, Nahmens, & Kennedy, 2003). For practical purposes, simulation sickness can be measured with the Simulation Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993), and further discussion of the issue can be found here (Oculus, 2017).

The challenge of presence

Before putting on a VR headset, it seems impossible to belief that a VR world could seem like the real thing, and indeed some VR scenarios are much more believable and 'real' than others. The term '*presence*' is often used to describe and evaluate the experience of VR making you feel like you are somewhere else (Sheridan, 1992; Usoh, Catena, Arman, & Slater, 2000), and '*co-presence*' or '*social presence*' is used to describe the experience of

being with someone else (Casanueva & Blake, 2001; Garau *et al.*, 2003). In 2009, Slater proposed the term '*place illusion*' for 'the strong illusion of being in a place in spite of the sure knowledge that you are not there'. In this context, the term '*immersion*' describes the technical ability of a system to support sensorimotor contingency, forming the framework in which place illusion could occur. The place illusion defines user's response to the system, taking into account the possibility that different people could have different experiences of the place illusion in the same system.

The same paper (Slater, 2009) also proposed to use the term '*plausibility illusion*' for the illusion that events happening in VR are real and that in order for the plausibility illusion to occur, the VR events should be relating personally to the user. This means the characters and items in the VR world respond to the user as the user interacts with the VR world (rather than just watching a 3D movie). For instance, a situation that triggers plausibility illusion could be when someone entering a virtual bar, a virtual character approaches them and starts a conversation (Pan *et al.*, 2012). In a typical setup of an experimental study using virtual characters in VR, the strength of the place illusion is often influenced by the VR display technology, whereas the strength of the plausibility illusion is influenced by the animation and interactivity of the virtual characters.

As both place and plausibility illusions are subjective concepts, a variety of questionnaires have been developed to measure them (Usoh *et al.*, 2000; Witmer & Singer, 1998). These ask questions such as 'To what extent did you have a sense that you were in the same place as person X?' and 'To what extend did you have a sense of being part of the group?' Individual differences in response could be caused by differences in personality, in multisensory integration (Haans, Kaiser, Bouwhuis, & IJsselsteijn, 2012), in prior experience of gaming/VR, or other factors. Further research will be needed to define these fully. A key point to note is that, while current VR systems may give a strong sense of place and presence, they still differ substantially from real life and no participants are confused between the two. Thus, for studying phenomena which rely on the belief that another person is present (e.g., the audience effect), it may be valuable to tell participants that a VC is actually driven by another person (even if it is not). Further, although virtual characters generally are perceived to be more plausible when they have more human-like animations and are programmed to be more interactive, higher level of graphical realism of those characters does not necessarily increase the level of co-presence.

The challenge of ethics

Psychology researchers have substantial experience in considering the ethical issues surrounding research, and studies with human participants are typically scrutinized by an ethics panel before data can be collected. VR research in psychology is subject to the same constraints, but some particular issues are worthy of examination. It is often suggested that one of VR's benefits is that it can be used to recreate dangerous or stressful situations to explore people's reaction, which would otherwise very difficult to study or even impossible. For instance, to test participant's fear responses, VR was used to create a 'room 101' with disturbing events such as spiders crawling around, explosions, and a floor collapsing (McCall *et al.*, 2015). Studies from the Slater group include a recreation of the famous Milgram experiment where participants had to execute fatal electric shots to a virtual character (Slater *et al.*, 2006); a violent fight scenario (Slater *et al.*, 2013) and a moral dilemma with an active shooter (Pan & Slater, 2011). In the latter, participants thought their task was to operate a lift in a gallery in VR, but later were shocked to be confronted with a tough decision: A gunman entered the lift and started shooting, and

within a few seconds they had to choose whether they should push a button to save five people but sacrificing one other, who would otherwise be fine.

It can be argued that putting participants through these scenarios in VR is the closest we can get to study their behaviours in a similar real-life situations, as participants react to virtual events and virtual characters as if they were real. Nevertheless, participants remain aware that there was no *real danger* nor were there *real consequences* (nobody is really hurt as a result of participants' decisions). However, as the VR hardware gets better in supporting 'immersion' and the virtual characters both appear and behave more and more realistic, the boundary between virtual and real is becoming blurrier. It is therefore particularly important to provide full information before participants take part, to make participants aware of their right to withdraw and to emphasise how to withdraw (e.g., close your eyes and say STOP to leave the virtual world and the experimenter will stop the study). Further, as various studies have showed that experiences in VR could lead to change in participants' behaviour and attitude in their real life (Banakou, Hanumanthu, & Slater, 2016; Tajadura-Jiménez, Banakou, Bianchi-Berthouze, & Slater, 2017) and can even create a false-memory in children (Segovia, Bailenson, Segovia, & Bailenson, 2009), the implication of VR experiences should be carefully discussed with participants.

A second potential ethical issue for studies in VR concerns personal disclosure, because some studies suggest that people may be more willing to disclose personal information (including abuse or trauma) to a virtual character than to a real person (Lucas, Gratch, King, & Morency, 2014; Rizzo *et al.*, 2015). This can be valuable in some treatment scenarios, but is also a risk. Confidential data collected from participants in VR could cause privacy concerns to a greater extent than data collected with traditional methods. For instance, in VR data collection often includes conversation exchange with avatars, gaze, and mocap data. These sensitive and personal information must be dealt with caution with relevant data protection measurements in place. More broadly, putting a person in a stressful situation to see how they behave could change their perception of themselves – for example, someone who finds themselves too panicked to help in a VR test of prosocial behaviour could potentially leave the study feeling like a 'bad person'. Fully informed consent and full debrief procedures may help here, but careful consideration of these issues and how to mitigate them is vital. For more detailed discussion of ethical issues around VR in gaming, research, and therapy contexts, we point the reader to Brey (1999) and Madary and Metzinger (2016).

Finally, despite all the ethical challenges, we must also not forget the great potential of VR to have a positive impact on our real life in various aspects, including science, education, medicine, and training. For more on this, we point the reader to Slater and Sanchez-Vives (2016).

The challenge of experimental design

Traditional studies in cognitive psychology or psychophysics may have participants perform the same type of trial dozens or hundreds of times over, to obtain precise measures of performance. In contrast, typical VR scenarios have a relatively short duration (it is hard to maintain presence over a long time) and participants might experience just one or two critical events. Thus, VR can call for very different kinds of experimental design. A further challenge arises in interactive VR, where a virtual character responds to the behaviour of a participant and the two take turns in a conversation. In such a scenario, each participant may experience a slightly different sequence of events, and it does not necessarily make sense to average all participants together. For example, in studies of

negotiation training (Gratch, Devault, *et al.*, 2016) and of bargaining (Gratch, Nazari, *et al.*, 2016), participants learn to negotiate with a virtual character but because the virtual character is responsive to the participant, each person will experience a slightly different set of offers in the game. This means different participants may reach different bargaining outcomes, depending on how they started the game and what decisions they made. Thus, data cannot necessarily be analysed by the typical method of averaging all participants together.

In some cases, it might make sense to treat the dyad (human + virtual character) as the unit of analysis, comparing how dyads reach one decision or another. An alternative may be to draw on research in neuroeconomics and develop a model of the human-VC behaviour in which the values and rewards that the human and VC assign to different options can be modelled on a trial-by-trial basis (Hampton *et al.*, 2008). However, for complex negotiations, there may not be suitable models available. A third option, applicable to non-verbal behaviour more than to negotiation, may be to develop different analysis strategies which capture specific patterns of action in the human-VC dyad. For example, wavelet coherent methods (Schmidt, Nie, Franco, & Richardson, 2014) and cross-recurrence methods (Dale & Fusaroli, 2014) have proven valuable in quantifying the behaviour of human–human dyads and might also be useful in modelling human-VC dyads. For all these approaches, more work will be needed to develop appropriate experimental designs and analysis methods, suitable for the study of dynamic interactions between humans and VCs.

The challenges (and benefits) for theory

Advances in psychology are often driven by the development of theories and the rigorous testing of these theories against experimental data. Here, VR can provide both a challenge and a benefit. VR challenges our theories because it requires a precise and well-specified theory which can be implemented in an artificial system. For example, a theory might suggest that mimicry leads to prosocial behaviour (Lakin, Jefferis, Cheng, & Chartrand, 2003), but to build mimicry into a VR system, we must answer much more detailed questions – how fast does mimicry occur? which actions are mimicked? how accurately etc.? By building a VR system which implements mimicry, we can begin to address these questions and test the theory in detail (Hale & Hamilton, 2016). Similarly, theories might suggest that joint attention is implemented in particular brain systems, but testing this required a VR implementation of joint attention (Schilbach *et al.*, 2010), which requires us to specify the duration of mutual gaze between the participant and VC, the timing of the looks to the object, and the contingencies between these behaviours. Thus, VR requires a precise and well-specified theory of the psychological processes under investigation.

More generally, the architecture of a 'virtual human' may have commonalities with our models of cognitive processing in real humans (Figure 3). Where a real person has a visual system, a virtual human must have machine vision and sensors to interpret the actions of their partner. Where a real person has a motor system, a virtual human must have a control policy to determine which actions to execute and when. And where a real person has brain systems for theory of mind, decision-making, affect sharing or reward processing, a virtual human might need to draw on similar systems. Just as research in machine vision and vision sciences can use similar or different computational models, so research into other aspects of human
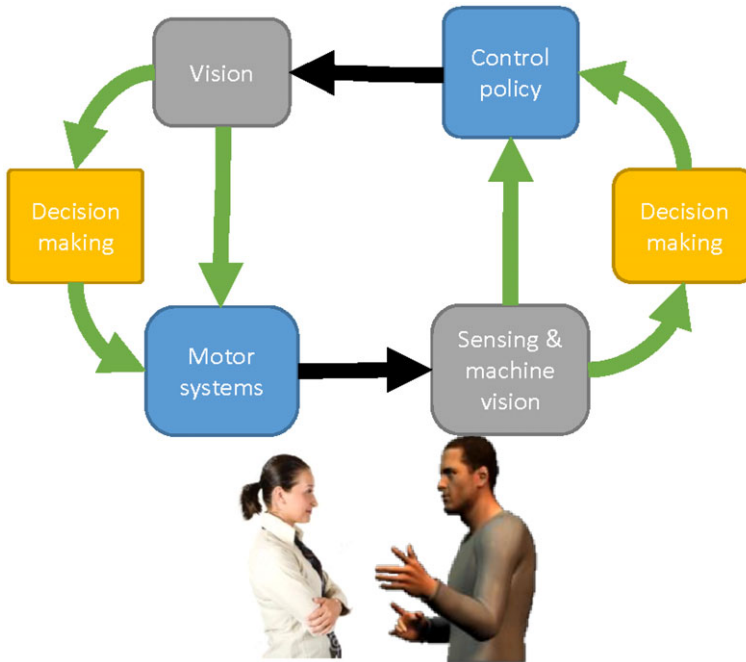
Figure 3. The human-virtual agent loop. Colour coding indicates how human cognitive processes have parallels in the control of virtual agents.

performance may (or may not) find parallels in the systems needed to implement realistic virtual behaviour. Finding where these parallels are and which are important will be a valuable endeavour.

To give a concrete example, Kopp and Bergman consider a number of possible cognitive models of speech and gesture production, before describing how these can be implemented in a virtual character. Examination of the behaviour of the character can be combined with simulations and behavioural data to both build better VR and test models of gesture control (Kopp & Bergmann, 2017). Thus, using VR imposes rigour on our psychological theories – a sloppy or weakly specified theory cannot be implemented in VR, whereas a precise theory will be able to guide the creation of good VR and can be tested at the same time. While this is a large challenge, achieving it will bring substantial benefits for the field. In particular, it represents an important step towards our final challenge.

Mons Olympus – the big challenge

As a guide to establishing a VR laboratory, our paper has thus far provided a map of the available technologies and a brief overview of common difficulties which can be avoided with care. We hope this outline will help researchers understand the practicalities of how to do VR research, but also highlight why one should (or should not) use VR in the study of human social behaviour. Knowing the boundary conditions of what a VR setup can achieve is critical in knowing where this technology can be of use. With these constraints in mind, the final section of this study examines the biggest

challenge in social VR research – the Mons Olympus of the field – building a fully interactive virtual human.

Imagining a VR Turing test

The original Turing test (Turing, 1950) was proposed as a way to determine if a computer has achieved human-like intelligence. It is typically implemented in a chatroom environment, where testers communicate with a person or a computer via the medium of text. The tester is asked to determine if the being in the other room is a human (pretending to be someone else) or a computer (pretending to be a human). Saygin, Cicekli, and Akman (2000) provide a fascinating history of the Turing test. In recent years, the Loebner prize contest has been held to compare chatbots which attempt to convince judges that they are human. A chatbot recently 'passed' this test, albeit using tricks including pretending to be a child from a different country, rather than by showing adult levels of behaviour (You, 2015). It is also possible for some non-verbal systems to effectively mimic the behaviour of humans. Participants in an interactive gaze study were asked to judge if their interaction partner was human or computer (Pfeiffer, Timmermans, Bente, Vogeley, & Schilbach, 2011), and their performance was close to chance, at least when they believed the person was attempting to deceive them.

Building on these, it is possible to imaging a VR version of the Turing test, in which testers determine if a virtual character has both human intelligence *and* human non-verbal behaviour. For example, participants meet a character in a VR space and must determine if that character is an avatar (controlled in real time by a human next door) or an agent (controlled entirely by a computer). Passing the VR Turing test seems at first glance to be much more challenging than a text-based Turing test. But how hard would it really be? In particular, could a VR Turing test be passed with a few hacks, putting together some previously recorded behaviours, maybe driven by some clever machine-learning algorithms, to trick users into believing they are interacting with a real person? Or, on the other hand, is the problem of passing the VR Turing test really a problem that is AI complete – that is – a problem which cannot be solved until we have placed the full intelligence of a real human into a computer system.

Part of the solution to this problem must lie in constraining *who* the VR system is attempting to emulate – it seems easier to emulate a person who is very unlike the judges or may have good reasons to not answer questions, than it would be to emulate a friend or colleague. Like the Eugene chatbot which passed a Turing test by emulating a Ukrainian boy (You, 2015), current and foreseeable VC systems can potentially do a good job of simulating human behaviour in a narrow field of knowledge and a narrow range of emotional expressiveness. But developing VC systems which demonstrate wider knowledge and more meaningful expressiveness will be valuable for theories of social cognition in two ways. First, we can consider which aspects of social behaviour can be implemented with simple, low level algorithms (tricks) and which require more complex processing of emotions or mental states. And second, we can dissect the algorithms which succeed in creating good VR characters to determine what makes them work.

Finally, if a believable VR character can be built, even for a limited field of knowledge, this would have enormous utility across a wide range of domains. Teaching and therapy are areas where VR characters are already being used but retail, customer service and business might also make use of these. It is for psychologists to make sure that our understanding of real human interactions keeps pace with the developments in artificial human interactions so that these two fields can gain maximum benefit from each other.

Conclusions

In this paper, we have lead the reader from a basic outline of the equipment needed for VR, across the landscape of possible experiments and glimpsed the future of virtual humans. We hope this target article will spark debate about the use of VR in psychology research and practice and act as a primer for researchers interested in exploring this exciting new domain.

Acknowledgements

References

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*(1), 32–47. https://doi.org/10.1016/j.newideapsych.2007.12.002

Bailenson, J. N., & Yee, N. (2007). Virtual interpersonal touch and digital chameleons. *Journal of Nonverbal Behavior*, *31*, 225–242. https://doi.org/10.1007/s10919-007-0034-6

Baltrusaitis, T., Robinson, P., & Morency, L.-P. (2016). *OpenFace: An open source facial behavior analysis toolkit*. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1–10). IEEE. https://doi.org/10.1109/wacv.2016.7477553

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 12846–12851. https://doi.org/10.1073/pnas.1306779110

Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in Human Neuroscience*, *10*, 601. https://doi.org/10.3389/fnhum.2016.00601

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*, 153–158. https://doi.org/10.1016/j.jesp.2016.02.003

Bohil, C. J., Alicea, B., & Biocca, F. a. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews. Neuroscience*, *12*, 752–762. https://doi.org/10.1038/nrn3122

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, *391*, 756. https://doi.org/10.1038/35784

Brey, P. (1999). The ethics of representation and action in virtual reality. *Ethics and Information Technology*, *1*, 5–14. Retrieved from https://link.springer.com/content/pdf/10.1023%2FA%3A1010069907461.pdf

Brooks, F. P. (1999). What's real about virtual reality? *IEEE Computer Graphics and Applications*, *19*(6), 16–27. https://doi.org/10.1109/38.799723

Cameron, J. (2009). *Avatar*. 20th Century Fox.

Casanueva, J., & Blake, E. H. (2001, January 19–28). *The effects of avatars on co-presence in a collaborative virtual environment*. Proc Ann Conf SA Inst of Computer Scientists and Information Technologists. Retrieved from https://www.cs.uct.ac.za/Members/edwin/publications/2001-copres.pdf

Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2001). *BEAT*. Proceedings of the 28th annual conference on Computer graphics and interactive techniques – SIGGRAPH '01 (pp. 477–486). ACM Press, New York, NY. https://doi.org/10.1145/383259.383315

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). *Surround-screen projection-based virtual reality*. Proceedings of the 20th annual conference on Computer graphics and interactive techniques – SIGGRAPH '93 (pp. 135–142). ACM Press, New York, NY. https://doi.org/10.1145/166117.166134

Dale, R., & Fusaroli, R. (2014). The self-organization of human interaction. *Psychology of Learning*, 59, 43–96. Retrieved from https://pure.au.dk/portal/files/53762658/2014_plm_the_self_organization_of_human_interaction.pdf

de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in Psychology*, 6, 576. https://doi.org/10.3389/fpsyg.2015.00576

de la Rosa, S., Ferstl, Y., & Bülthoff, H. H. (2016). Visual adaptation dominates bimodal visual-motor action adaptation. *Scientific Reports*, 6(1), 23829. https://doi.org/10.1038/srep23829

de la Rosa, S., Lubkull, M., Stephan, S., Saulton, A., Meilinger, T., Bülthoff, H., & Cañal-Bruland, R. (2015). Motor planning and control: Humans interact faster with a human than a robot avatar. *Journal of Vision*, 15(12), 52. https://doi.org/10.1167/15.12.52

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., . . . Morency, L.-P. (2014). *SimSensei Kiosk: A virtual human interviewer for healthcare decision support*. International Conference on Autonomous Agents and Multi-Agent Systems (pp.1061–1068). Retrieved from https://dl.acm.org/citation.cfm?id=2617415

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE*, 7(1), e29081. https://doi.org/10.1371/journal.pone.0029081

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guide-lines for research and an integration of findings*. Pergamon Press.

Eyben, F., Wöllmer, M., & Schuller, B. (2010). *Opensmile*. Proceedings of the international conference on Multimedia – MM '10 (pp. 1459). ACM Press, New York, NY. https://doi.org/10.1145/1873951.1874246

Ferstl, Y., Bülthoff, H., & de la Rosa, S. (2017). Action recognition is sensitive to the identity of the actor. *Cognition*, 166, 201–206. https://doi.org/10.1016/j.cognition.2017.05.036

Forbes, P. A. G., Pan, X., & Hamilton, A. F. de. C. (2016). Reduced mimicry to virtual reality avatars in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 46, 1–10. https://doi.org/10.1007/s10803-016-2930-2

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). *The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment*. Proceedings of the conference on Human factors in computing systems – CHI '03 (p. 529). ACM Press, New York, NY. https://doi.org/10.1145/642611.642703

Gratch, J., Devault, D., & Lucas, G. (2016). The benefits of virtual humans for teaching negotiation. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 10011 LNAI, pp. 283–294). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-47665-0_25

Gratch, J., & Hartholt, A. (2013). Virtual humans: A new toolkit for cognitive science research. *McNeill*, 2005, 41–42. Retrieved from https://pdfs.semanticscholar.org/a195/159e52ad6708897012f02baa469f72830e3c.pdf

Gratch, J., Nazari, Z., & Johnson, E. (2016). *The misrepresentation game: How to win at negotiation while seeming like a nice guy*. AAMAS 2016, (Aamas), (pp. 728–737). Retrieved from http://trust.sce.ntu.edu.sg/aamas16/pdfs/p728.pdf

Haans, A., Kaiser, F. G., Bouwhuis, D. G., & IJsselsteijn, W. A. (2012). Individual differences in the rubber-hand illusion: Predicting self-reports of people's personal experiences. *Acta Psychologica*, *141*, 169–177. https://doi.org/10.1016/j.actpsy.2012.07.016

Hale, J., & Hamilton, A. F. de. C. (2016). Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific Reports*, *6*, 35295. https://doi.org/10.1038/srep35295

Hale, J., Payne, M. E. M., Taylor, K. M., Paoletti, D., & Hamilton, A. F. de. C. (2017). The virtual maze: A behavioural tool for measuring trust. *The Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2017.1307865

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6741–6746. https://doi.org/10.1073/pnas.0711099105

Hasler, B. S., Hirschberger, G., Shani-Sherman, T., & Friedman, D. A. (2014). Virtual peacemakers: Mimicry increases empathy in simulated contact with virtual outgroup members. *Cyberpsychology, Behavior and Social Networking*, *17*, 766–771. https://doi.org/10.1089/cyber.2014.0213

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology*, *25*, R621–R634. https://doi.org/10.1016/j.cub.2015.05.052

Kahl, S., & Kopp, S. (2017). *Self-other distinction in the motor system during social interaction: A computational model based on predictive processing*. Proceedings of the 39th Annual Conference of the Cognitive Science Society. Retrieved from https://pub.uni-bielefeld.de/download/2910276/2911411

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*, 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Kopp, S., & Bergmann, K. (2017). Using cognitive models to understand multimodal processes: The case for speech and gesture production. In S. Oviatt, B. Schuller & P. Cohen (Eds.), *The handbook of multimodal-multisensor interfaces: Foundations, user modeling, and common modality combinations* (Vol. 1, pp. 239–276). New York, NY: ACM. https://doi.org/10.1145/3015783.3015791

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72. https://doi.org/10.3758/s13423-012-0341-8

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*, 145–162. https://doi.org/10.1023/1025389814290

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

MacDorman, K. (2006, January 26–29). *Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley*. ICCS/CogSci-2006 Long Symposium: Toward social mechanisms of android science. https://doi.org/10.1093/scan/nsr025

Madary, M., & Metzinger, T. K. (2016). Recommendations for good scientific practice and the consumers of VR-technology. *Frontiers in Robotics and AI*, *3*, 3. https://doi.org/10.3389/frobt.2016.00003

Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2014). Changing bodies changes minds: Owning another body affects social cognition. *Trends in Cognitive Sciences*, *19*, 6–12. https://doi.org/10.1016/j.tics.2014.11.001

Mareschal, I., Calder, A. J., & Clifford, C. W. G. (2013). Humans have an expectation that gaze is directed toward them. *Current Biology*, *23*, 717–721. https://doi.org/10.1016/j.cub.2013.03.030

McCall, C., Hildebrandt, L. K., Bornemann, B., & Singer, T. (2015). Physiophenomenology in retrospect: Memory reliably reflects physiological arousal during a prior threatening experience. *Consciousness and Cognition*, *38*, 60–70. https://doi.org/10.1016/j.concog.2015.09.011

McCall, C., & Singer, T. (2015). Facing off with unfair others: Introducing proxemic imaging as an implicit measure of approach and avoidance during social interaction. *PLoS One*, *10*(2), e0117532. https://doi.org/10.1371/journal.pone.0117532

Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Oculus. (2017). *Simulator sickness*. Retrieved from https://developer.oculus.com/design/latest/concepts/bp_app_simulator_sickness/

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pan, X., Gillies, M., Barker, C., Clark, D. M., & Slater, M. (2012). Socially anxious and confident men interact with a forward virtual woman: An experimental study. *PLoS One*, *7*, e32931. https://doi.org/10.1371/journal.pone.0032931

Pan, X., & Hamilton, A. F. de. C. (2015). Automatic imitation in a rich social context with virtual characters. *Frontiers in Psychology*, *6*, 790. https://doi.org/10.3389/fpsyg.2015.00790

Pan, X., & Slater, M. (2011). *Confronting a moral dilemma in virtual reality: A pilot study*. Proceedings of the 25th BCS Conference on Human-Computer Interaction. British Computer Society. Retrieved from https://dl.acm.org/citation.cfm?id=2305326

Patton, J., Dawe, G., Scharver, C., Mussa-Ivaldi, F., & Kenyon, R. (2006). Robotics and virtual reality: A perfect marriage for motor control research and rehabilitation. *Assistive Technology*, *18*, 181–195. https://doi.org/10.1080/10400435.2006.10131917

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, *22*, 779–787. https://doi.org/10.1016/j.concog.2013.04.016

Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass: Processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science*, *15*, 598–603. https://doi.org/10.1111/j.0956-7976.2004.00726.x

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLoS One*, *3*(12), e3832. https://doi.org/10.1371/journal.pone.0003832

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., & Schilbach, L. (2011). A non-verbal turing test: Differentiating mind from machine in gaze-based social interaction. *PLoS One*, *6*(11), e27591. https://doi.org/10.1371/journal.pone.0027591

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience and Biobehavioral Reviews*, *37*(10 Pt 2), 2516–2528. https://doi.org/10.1016/j.neubiorev.2013.07.017

Pine, D. S., Grun, J., Maguire, E. A., Burgess, N., Zarahn, E., Koda, V., . . . Bilder, R. M. (2002). Neurodevelopmental aspects of spatial navigation: A virtual reality fMRI study. *NeuroImage*, *15*, 396–406. https://doi.org/10.1006/nimg.2001.0988

Piwek, L., McKay, L. S., & Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, *130*, 271–277. https://doi.org/10.1016/j.cognition.2013.11.001

Rizzo, A., Cukor, J., Gerardi, M., Alley, S., Reist, C., Roy, M., . . . Difede, J. (2015). Virtual reality exposure for PTSD due to military combat and terrorist attacks. *Journal of Contemporary Psychotherapy*, *45*, 255–264. https://doi.org/10.1007/s10879-015-9306-3

Rose, F. D., Brooks, B. M., & Rizzo, A. A. (2005). Virtual reality in brain damage rehabilitation: Review. *CyberPsychology & Behavior*, *8*, 241–262. https://doi.org/10.1089/cpb.2005.8.241

Sacheli, L. M., Christensen, A., Giese, M. A., Taubert, N., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2015). Prejudiced interactions: Implicit racial bias reduces predictive simulation during joint action with an out-group avatar. *Scientific Reports*, *5*, 8507. https://doi.org/10.1038/srep08507

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. D. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7, 413–422. https://doi.org/10.1093/scan/nsr025

Saygin, A. P., Cicekli, I., & Akman, V. (2000). Turing test: 50 years later. *Minds and Machines*, *10*, 463–518. https://doi.org/10.1023/A:1011288000451

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., . . . Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, *22*, 2702–2715. https://doi.org/10.1162/jocn.2009.21401

Schmidt, R. C., Nie, L., Franco, A., & Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Frontiers in Human Neuroscience*, 8, 633. https://doi.org/10.3389/fnhum.2014.00633

Schröder, M. (2012). *The SEMAINE API: A component integration framework for a naturally interacting and emotionally competent embodied conversational agent*. Retrieved from http://scidok.sulb.uni-saarland.de/volltexte/2012/4544/

Segovia, K. Y., Bailenson, J. N., Segovia, K. Y., & Bailenson, J. N. (2009). Virtually true: Children's acquisition of false memories in virtual reality. *Media Psychology*, *12*, 371–393. https://doi.org/10.1080/15213260903287267

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, *1*(1), 120–126. https://doi.org/10.1162/pres.1992.1.1.120

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1535), 3549–3557. Retrieved from http://rstb.royalsocietypublishing.org/content/364/1535/3549.short

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., . . . Sanchez-Vives, M. V. (2006). A virtual reprise of the stanley milgram obedience experiments. *PLoS One*, *1*(1), e39. https://doi.org/10.1371/journal.pone.0000039

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2008). Towards a digital body: The virtual arm illusion. *Frontiers in Human Neuroscience*, 2, 6. https://doi.org/10.3389/neuro.09.006.2008

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience*, *3*, 214–220. https://doi.org/10.3389/neuro.01.029.2009

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., & Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLoS One*, *8*(1), e52766. https://doi.org/10.1371/journal.pone.0052766

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74. https://doi.org/10.3389/frobt.2016.00074

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS One*, *5*(5), e10564. https://doi.org/10.1371/journal.pone.0010564

Spanlang, B., Normand, J.-M., Borland, D., Kilteni, K., Giannopoulos, E., Pomés, A. s., . . . Slater, M. (2014). How to build an embodiment lab: Achieving body representation illusions in virtual reality. *Frontiers in Robotics and AI*, *1*, 9. https://doi.org/10.3389/frobt.2014.00009

Stanney, K. M., Hale, K. S., Nahmens, I., & Kennedy, R. S. (2003). What to expect from immersive virtual environment exposure: Influences of gender, body mass index, and past experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*, 504–520. https://doi.org/10.1518/hfes.45.3.504.27254

Steed, A., Pan, Y., Zisch, F., & Steptoe, W. (2016). The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)* (pp. 67–76). IEEE. https://doi.org/10.1109/vr.2016.7504689

Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., & Slater, M. (2017). Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Scientific Reports*, *7*(1), 9637. https://doi.org/10.1038/s41598-017-09497-3

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Tsakiris, M. (2008). Looking for myself: Current multisensory input alters self-face recognition. *PLoS One*, *3*(12), e4040. https://doi.org/10.1371/journal.pone.0004040

Turing, A. M. (1950). Turing+mind+1950.pdf. *Mind*, *59*, 433–460. https://doi.org/10.2307/2251299

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, *9*, 497–503. https://doi.org/10.1162/105474600566989

Verberne, F. M. F., Ham, J., Ponnada, A., & Midden, C. J. H. (2013). Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust. In *Proceedings of the 8th international conference on persuasive technology*, Sydney, NSW, Australia, April 03–05, 2013 (pp. 234–245). Heidelberg, Germany: Springer-Verlag Berlin. https://doi.org/10.1007/978-3-642-37157-8_28

Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, *38*(1), 174–180. https://doi.org/10.3758/BF03192765

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7, 225–240. https://doi.org/10.1162/105474698565686

Yoshida, W., Seymour, B., Friston, K. J., & Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*, 10744–10751. https://doi.org/10.1523/JNEUROSCI.5895-09.2010

You, J. (2015). Beyond the turing test. *Science*, *347*(6218), 116. https://doi.org/10.1126/science.347.6218.116

Zanon, M., Novembre, G., Zangrando, N., Chittaro, L., & Silani, G. (2014). Brain activity and prosocial behavior in a simulated life-threatening situation. *NeuroImage*, *98*, 134–146. https://doi.org/10.1016/j.neuroimage.2014.04.053

Zhao, R., Sinha, T., Black, A. W., & Cassell, J. (2016). *Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavior* (pp. 218–233). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-47665-0_20

Supporting Information

The following supporting information may be found in the online edition of the article:

Appendix S1. Summary of current VR display systems.
Appendix S2. Listing of some current mocap technologies.
Appendix S3. Listing of some current software providers and resources.

Why and how to use virtual reality to study human social interaction:
the challenges of exploring a new research landscape

Appendixes & Technical details

There are a number of technical solutions used to display images and to track the body in VR. Here we provide here a little more detail on the differences between these systems and some issues that must be considered in their use. We list display systems and their properties in Appendix 1. We list motion capture systems and compare them in Appendix 2. We list some common software packages in Appendix 3. All information is current at summer 2017 but this is a rapidly developing field and new systems are likely to become available.

Appendix 1 – Summary of current VR display systems

Table 1: summary of VR systems This table compares different VR systems (CAVE, HMDs, Projector, TV) often used in psychology studies.

	3D stereo	Field of view	Head-tracking	Category
CAVE	Yes	Max width as natural eye But not 360 (usually 4 walls with no back or ceiling wall)	Yes, both position and rotation tracking	Immersive VR
HMD	Yes	Differs from devices to device Full surrounding no escape	Yes, some both, some rotation tracking only	Immersive VR
Projector	Possible (3D glasses)	Depends on screen size and user distance One wall	Not with most systems	Semi-immersive VR
Monitor/TV	Possible (3D glasses)	Depends on screen size and user distance	No	Non-immersive VR

Table 2: Summary of Technical features of HMD devices (as of July 2017)

Category	Tracking	Name	Resolution (pre-eye)	Refresh rate	Weight	Controller
Mobile VR	3DoF	Samsung Gear VR	1280 x 1440*	60 Hz*	318g + phone	some are 3Dof
Console VR	6DoF (external camera)	PlayStation VR	960 x 1080	120 Hz	610g	6DoF tracked vibration
Desktop HMDs	6DoF (external camera/light house)	Oculus Rift CV1	1080 x 1200	90 Hz	470g	6DoF tracked vibration
		HTC VIVE	1080 x 1200	90 Hz	555g	6DoF tracked vibration

* depending on the mobile device

Appendix 2. Listing of some current mocap technologies

(note there are many other systems which are not listed here)

Company & Device	What does it track?	How does it track?	Typical spec
Faceware	Facial action units	Helmet-mounted camera	30Hz
Leapmotion	Hands	3D camera (markerless)	Up to 1mm
Metria	Body / hands	Moire phase markers + camera	Up to 0.2mm and orientation data
Microsoft Kinect	Whole body motion	3D camera (markerless)	30Hz and up to 1-40mm
Optitrack (NaturalPoint)	Body / face / hands	Reflective markers + cameras	Up to 360Hz Approx. 1mm
Optotrack (northernDigital)	Hands / body / face	Small wired LED markers	Up to 4600Hz and 0.1mm
PerceptionNeuron	Whole body	Inertia / magnetic units (can be wireless)	Up to 120fps; accuracy not specified
PhaseSpace	Whole body	Reflective markers + cameras	Up to 960Hz and under 1mm
Polhemus	Hands / body	Wired magnetic markers	Up to 120Hz and <1mm accuracy, plus marker orientation
Synertial	Hands / body	Inertia units in glove / suit	Up to 500Hz; accuracy not specified
Vicon	Body / face / hands	Reflective markers + cameras	Up to 420 Hz Approx 1mm

Commentary on tracking the body & representing the self in VR

As listed above, there are a wide range of different mocap systems available which use very different methods to track the body. Markerless systems (e.g. Kinect) use RGB and infrared (IR) cameras together with advanced image processing to infer the body location. Passive marker system (Vicon / Optitrak) use multiple IR cameras to track reflective markers on key joints, calculating the 3D position of each marker (3DoF). Other solutions involve active markers (e.g. magnetic markers such as Polhemus), which tracks 6DoF data (position and rotation) in real-time, or inertial measurement units (e.g., Perception Neuron).

To understand the advantages & disadvantages of each, it is helpful to understand how mocap data is typically used. In many studies, the raw signals from the mocap sensors are be used to create a skeleton representation of a moving participant, which is very useful for standard analyses and rendering. To do this, the mocap software must solve the inverse kinematics problem

– that is – it must map signals related to the location of markers on the limbs onto a 3D model of the skeleton in space which maintains consistent limb lengths and plausible joint angles as it moves.   For just a single marker on the hand, the inverse kinematics problem is unsolvable because many possible arm / elbow angles could hold the hand at the same location, and even with more markers, the mapping of raw data to a skeleton may be unreliable. To improve data quality, it helps to increase the number of markers, get more from the markers (6DoF rather than 3DoF), and to customise the limb-lengths of the skeleton to the person being captured. Different solutions to the inverse kinematics problem are built into different VR software and mocap packages, but may not always be easy to apply.

Whether building a skeleton model or just tracking individual markers, different tracking technologies come with their own pros and cons. Inertial sensors, or IMUs (Inertial Measurement Units), are portable as they are relatively small in size (as commonly integrated in smart phones). They work well on a short time scale but are prone to drifting over longer period of time. Thus, for social interaction research in the area of how two people coordinate their limb movements, inertial tracking could be problematic. When it comes to optical tracking, range of capture and occlusion are the two biggest challenges. Range of capture refers to the 3D space can be captured by the camera system, which is determined by the camera setup. Occlusion is the problem that one moving object (e.g. a participant's hand) may move in a way that occludes another object (e.g. a participant's head) from the view of the camera system, which means tracking data is lost. Systems with multiple cameras can reduce this problem but cannot always eliminate it. Magnetic systems (e.g. Polhemus) are robust to occlusion but are sensitive to electromagnetic interference and are best used in spaces which are free of metal and electronics, which is not always easy to achieve in a busy lab.   One major advantage of using high resolution body motion capture is to allow a detailed analysis of motion kinematics (Krishnan-Barman, Forbes, & Hamilton, 2017), but not all mocap systems provide sufficient accuracy for these analyses. In particular, most markerless systems and inertia systems (Kinect, LeapMotion & Perception Neuron) do not have the spatial and temporal resolution required for detailed kinematic analyses.

For all systems, it is also critical to consider the issue of latency, sampling frequency, and framerate. The latency between detecting a motion and updating the VR display accordingly can be a critical factor in building a successful VR application. For example, if a mocap system has a sampling rate of 60Hz (approx. 16 msec delay), the software has a delay of 20 msec to update the 3D models and animation, and the HMD itself has a refresh rate of 90Hz (approx. 12 msec delay) to update the virtual cameras and render the graphics according to the user's viewpoint, then the total system delay of 48 msec may be enough to disrupt the feeling of self-embodiment. In social VR scenarios, delay is less of an issue because latencies in typical social interactions are substantially longer than 60 msec; but for self-embodiment, it is critical to build a system with minimal delays. High frequency mocap systems and more optimised code can reduce latencies to acceptable levels for most applications, but this remains an important issue to consider. Finally, if external tracking devices (rather than the build-in HMD tracker) are used to track the user's viewpoint for VR rendering, one has to be very careful as this could slow down framerate to below 60Hz and cause simulation sickness.

Overall, there are a number of mocap systems available which vary substantially in accuracy and price, and different solutions will be appropriate to different studies. A combination of different

tracking technologies could be used together; however, this often leads to other technical challenges such as interference and system integration. Maintaining a well-integrated VR system with various tracking solutions could be time consuming and requires enthusiasm for technology, which remains a challenge for many psychologists. A dedicated technician in VR systems is often needed to support social interaction research labs using VR.

Appendix 3. Listing of some current software providers and resources

Name	Functionality	Characteristics
Unity3D	VR/Game engine	Scripting language: C# and JavaScript. Big online community, many plugins.
Worldviz Vizard	VR/Game engine	Scripting language: Python. Easy to learn
Unreal Engine	VR/Game engine	Scripting language: C++. High-end graphics for realism
A-Frame	Web framework for VR experience	Scripting language: JavaScript. Web based VR.
Autodesk 3DsMax	Modelling and animation tool	3D modelling and animation
Autodesk MotionBuilder	Animation tool	3D character animation software
Autodesk Maya	Modelling and animation tool	3D modelling and animation
Z-brush	Modelling tool	Digital sculpting tool
Poser	Modelling tool	Specially for human figures
Mixamo	3D characters tool	Specially for human figures
Virtual Human Toolkit (USC)	Animate agent with face / body behaviour	System for creating virtual agents https://vhtoolkit.ict.usc.edu/
OpenFace	Analysis of facial behaviour	https://www.cl.cam.ac.uk/research/rainbow/projects/openface/
Behavior Expression Animation Toolkit (BEAT)	Synthesise speech & nonverbal behaviour	https://dl.acm.org/citation.cfm?id=383315
IAGO	Platform for online negotiation with virtual agents	http://people.ict.usc.edu/~mell/IAGO/