

# Removing the mask - do people over trust avatars reconstructed from video?

Simon Champion<sup>1</sup>, Aleksandra Landowska<sup>2</sup>, Tobias Duckworth<sup>3</sup>, Carl Moore<sup>4</sup>, Peter Eachus<sup>5</sup> and David Roberts<sup>6</sup>

<sup>1,2,3,4,5,6</sup> The University of Salford, Salford M5 4WT, UK  
s.p.campion@salford.ac.uk

**Abstract.** This experiment compared the detection of deceit across video conferencing and a fixed viewpoint 3D video based computer graphic medium. The purpose was to determine if the process of 3D reconstruction influenced trust by reducing detail of facial expression. Comparison with the literature investigates the impact of facial expression on trust. Inspiration comes from previous studies in the natural and virtual world that suggest a stronger tendency to over trust a person when their facial expression is hidden. A virtual avatar that copies head and eye movement but not that of the face, could be argued as akin to a person wearing a mask. Thus, our opening research question is: Would a 3D medium that removed this mask result in a truth bias similar to video and therefore real world? Two confederates each gave a set of accounts of which half were true. These were captured and transmitted simultaneously in real time using 2D and full 3D video based communication mediums. Recordings of these sessions were later examined by two sets of participants. Twenty-one participants were asked to determine which accounts were true. Measures included: accuracy at detecting truth and deceit, and from this tendency to over trust and lastly cognitive effort in determining truthfulness. Results show that participants performed and worked to a similar degree in both mediums. Findings are of interest to those developing 3D telepresence technologies and virtual humans, and to those concerned with the trustworthiness of a medium.

**Keywords:** Telepresence, Virtuality, Avatar, Video conferencing

## 1 Introduction

This paper presents a study that compared the detection of deceit across video conferencing and a 3D video based computer graphic medium. The inspiration for this experiment came from a small number of previous studies in the natural [1] and virtual world [2, 3] that suggest a stronger tendency to over trust a person when their facial expression is hidden. However, due to the small number of studies that have been undertaken, more work would be needed in this area before firm conclusions could be made. A virtual avatar that copies head and eye movement but not face movements, could be argued as being akin to a person wearing a mask. Thus the research question is: Would a 3D medium that removed this mask, result in a truth bias akin to video and

thus the real world? Two confederates (senders) each gave a set of accounts, of which half were true. These were captured and transmitted simultaneously using 2D and full 3D video based communication mediums. Recordings of these sessions were later examined by a set of participants. N=21 participants (receivers) who were asked to determine which accounts were true. Measures were taken of: accuracy at detecting truth and deceit, tendency to over trust and the effort undertaken in determining truthfulness. Data were captured using a questionnaire and functional near infrared spectroscopy. The results showed that, despite an obvious visual reduction in quality, the receivers performed and worked to a similar degree in both mediums. The findings of this experiment are of interest to those developing 3D telepresence technologies and virtual humans, and to those concerned with the trustworthiness of a medium.

Previous studies in virtual reality have discussed avatars that hide true identity and appearance [4]. Such avatars could have an application in protecting anonymity and reducing prejudice. However, identity and appearance are relevant in many situations where truthfulness must be determined. Furthermore, spatial aspects of non-verbal behaviour, such as mutual eye gaze, are relevant to the detection of deceit [5], thus giving a reason to use 3D other than to hide identity. The researcher wanted to know if a 3D avatar, designed to faithfully portray identity and appearance, would have sufficient quality of facial reproduction to avoid an exaggeration of the truth bias. Any method for capturing and reproducing a person in 3D induces errors. Previous truth detection studies have limited these errors by using camera tracking to accurately detect markers on the face, and doing much of the modelling off line. Approaches to faithfully reproducing appearance and identity in real time, especially when not impacting on the naturalness of behaviour, can ill afford such luxuries.

In essence, representation through a pre-modelled avatar, brought to life by motion tracking, could be argued as being akin to wearing a mask. The overarching research question is: would a 3D medium, that removed this mask, result in a truth bias akin to video and thus the real world? In order to faithfully communicate identity, appearance and attention, we are developing a telepresence approach that uses shape-from-silhouette to reconstruct users from live video allowing people to walk around 3d reconstructed versions of each other. Past studies [6] illustrate how this approach impacts on facial reproduction, and suggest it offers a direct comparison to video. When describing the technology approach used, the terms virtuality avatar and virtuality telepresence are used to distinguish from conventional authored avatars and the immersive collaborative virtual environments that use them. The term virtuality is used as these avatars reflect the current real appearance of their owners, in addition to simplifying their movements. A conventional avatar is authored offline and later driven by the tracked motion of its owner. As discrete tracking points drive the simulation of movement, some compare this to a puppet following strings.

Throughout this work the term sender will be used to represent the person communicating (truth or lies) and the term receiver for the person detecting the truthfulness of the communication. These terms are commonly used in the psychology literature on lie detection.

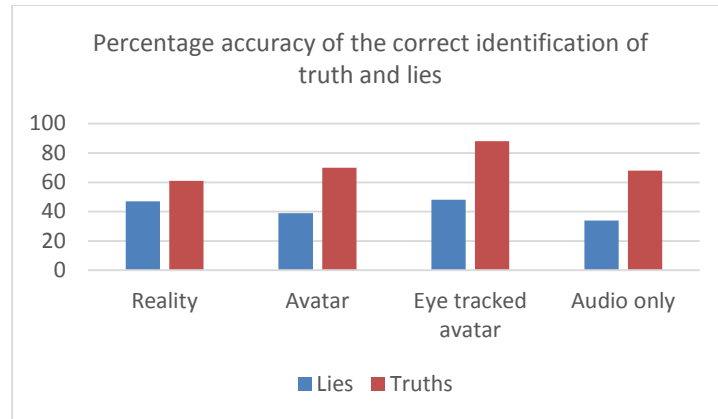
## 2 Background

### 2.1 Detection of truths and the fidelity of facial expression

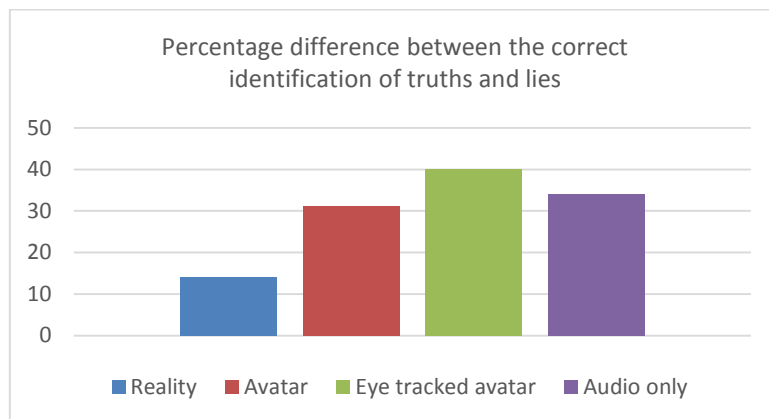
Detecting deception is of great importance across several fields including within judicial, security and law enforcement services. In studies using law enforcement personnel, accuracy rates when detecting truths range from 45-60% [7], and media quality has been demonstrated to significantly impact upon detection rates [8]. It has been shown that training in deception detection can significantly increase accuracy [9] and it has been suggested that virtual environments may be suited to deception detection training [10]. Research has been conducted by psychologists into the identification of deception using facial expressions [11-13] and VR groups have studied lie tracking using eye-tracked avatars [2, 3].

### 2.2 Accuracy of determining truthfulness in the natural and virtual world

When an individual has a tendency to over trust/believe someone is telling the truth regardless of a statement's veracity, this is referred to as truth bias [14]. A comparison of the literature suggests that during lie detection people tend to over trust others in both the real [1] and virtual world [2, 3] and over trust more when facial expression is not represented [2]. The detection of deception without any training is a difficult task and most people are poor at it. In an extensive meta-analysis Bond and DePaulo [1] synthesised the results of 206 studies involving 24,483 receivers. This analysis reported an overall accuracy of 54% yet, interestingly, there was a bias towards believing people were telling the truth, regardless of the truthfulness of their statement. The researcher has found only two studies using virtual avatars that present data comparable with Bond and DePaulo's study [1]. Both studies used motion tracking to drive head and eye movement. The study that additionally used motion tracking to drive facial expression [3] reported a truth bias comparable with that of the study by Bond and DePaulo [1]. The study that did not use motion tracking to drive facial expression [2] reported an exaggerated tendency to trust. While a comparison of these studies suggests that facial expression impacts on truth bias, there is an insufficient number of studies to draw a definitive conclusion. In Graph 1, the correct identification of truth and lies from Bond and DePaulo [1] and Steptoe's studies are shown [2], while in Graph 2 the percentage difference between the correct identification of truth and lies in these studies is illustrated. Unfortunately, the researcher could not directly contrast these results with the work of McDonnell et al. [3] because accuracy breakdowns were not available, however, McDonnell et al. did report a normal truth bias similar to that shown in previous psychological literature. Graph 2 shows that, in reality, a difference of around 14% between the identification accuracy rates of truth and lies can be expected. However, avatar studies (31% and 40%) are producing figures closer to audio only communication (34%). Could this be due to lack of facial expressions being communicated? Due to the small number of studies identified and varying experiment designs a firm conclusion cannot be drawn, however, it is interesting and more studies utilising avatars are needed to investigate this.



**Graph 1.** The percentage accuracy of the correct identification of truth and lies. Reality values taken from Bond and DePaulo's meta-analysis [1] and all other values taken from the study conducted by Steptoe et al. [2].



**Graph 2.** The percentage difference between the correct identification of truth and lies. Reality values taken from Bond and DePaulo's meta-analysis [1] And all other values taken from the study conducted by Steptoe et al. [2].

### 2.3 Measurement

Both the above studies measured performance in terms of the accuracy of determining truthfulness, split between true and false statements. To tie this to the impact of quality of facial expression, it would be interesting to quantify what facial expressions each medium could show. To strengthen this link it would be useful to know how hard people had to work to detect deceit in both mediums. The section below summarises related work in the identification of facial expressions and the methods of measuring cognitive workload.

Facial expressions, it is suggested, are in-built human behaviours which occur in all cultures [15]. Micro expressions are facial expressions that occur very quickly. Humans cannot control them and they happen in a split second, leaking information about feelings [11]. Previous studies on micro expressions show that they may allow medical doctors to better understand the needs of patients by identifying emotions such as fear or sadness [16]. With the push towards telehealth, a practitioner's ability to identify patient emotions through media could be important. Micro expression tuition is being researched as a way of helping military and law enforcement personnel to detect deception [17, 18]. Micro expressions may also have applications in the judicial system [19]. A facial coding system has been created to map and identify micro expressions [20] and previous studies have investigated how accurately facial expressions and emotions can be identified using avatars [21].

The purpose of mental workload measurement is to assess the effect of cognitive demands that are associated with task performance. Previous studies have demonstrated that high cognitive workload is positively correlated with greater neuronal activation in the prefrontal cortex [22], as measured by functional near-infrared spectroscopy (fNIRS). fNIRS is a non-invasive optical neuroimaging method for mapping the functioning of the human brain. It offers better spatial resolution than EEG and more temporal resolution than fMRI (functional magnetic resonance imaging) [23]. fNIRS is also, arguably, less likely to influence results than fMRI due to its low encumbrance, the fact that it does not enclose the subject, and its quick set up. These factors make it a promising neuroimaging technique for investigating cognitive workload in virtual environments and, indeed, it has already been used in previous studies [24-26].

#### **2.4 Approaches to capturing the face and reproducing it in 3D**

Approaches to capturing facial expressions for authored avatars either track specific parts of the face such as eyes and mouth, or the face as a whole. The former is well suited to driving an avatar face that is visually different to that of the owner. For the latter, a model of facial muscles within the avatar is driven by motion tracking data. Both marker based [27] and markerless [28] capture approaches are popular. Marker based tracking is more accurate, whereas markerless tracking does not require large numbers of tracking markers stuck to the face. The ability to accurately track the muscle groups of the face is significant as a smile does not only affect the mouth, it also engages muscle groups near the eyes [18]. Without the muscles of the face being mimicked it might be more difficult for viewers to identify emotions. However, driving the simulation of muscle movements is not the only approach to mimicking facial expression.

Other approaches, such as those used for telepresence [29, 30] derive a fluid 3D reconstructed avatar from live video. These approaches attempt to faithfully reproduce appearance, not just movement. Popular 3D reconstruction approaches in this area are shape-from-silhouette [31] and depth-based [32, 33]. Both impact on the quality of reproduction in different ways. Shape-from-silhouette cannot model concavities, such as eye sockets (although texture information can simulate this). However, when cameras are satisfactorily calibrated together, the visual-spatial quality increases with the number of cameras. The predominant depth-based approach [32, 33] bounces infrared structured light off the subject. The resolution of the depth-map is thus determined by the resolution of the structure reflecting off the subject rather than that of the camera.

Depth-based cameras are also used to drive pre-modelled avatars, including the mimicking of facial expression [28].

Abstract avatar approaches can hide identity and details of appearance. Such hidden details include perspiration and small facial movements which are not reflected by marker driven models of muscles, although change in skin tone (blushing) has been modelled [34]. All these details can be captured using 3D reconstruction approaches. Abstract avatars are suited to applications where protecting anonymity is important, thus increasing a person's confidence and perhaps giving them an unprejudiced hearing. In the real world a person's physical appearance can impact upon how honest one perceives them to be [35]. It should, therefore, be considered that avatar representation can impact upon trust. 3D reconstructed avatars are suited to giving an accurate portrayal at the expense of anonymity. They could, for example, be well suited to virtual identification parades [36]. This has the advantage of allowing witnesses to identify criminals without being present and the task could take place in a virtual crime scene, perhaps therefore helping with recollection.

### **2.5 Validity of comparing video based 3D reconstruction to video.**

The aforementioned meta study of the detection of deception [1] includes many studies that have used video to ensure that participants are given consistent data. High quality video captures the vast majority of facial expressions. Facial action coding [20] typically uses video both for repeatability and for frame by frame analysis. 3D reconstructed video provides a fairer comparison with video conferencing, than with an abstract virtual medium, as all modelling is live. A medium can be more easily compared to video if its contents are derived from that video and attempts a faithful reproduction. Firstly, exactly the same input data can be used for 2D and 3D video, albeit only one channel can be used for 2D video. Secondly, any visual difference is an unwanted consequence of the reconstruction process rather than a wanted abstraction. These reconstruction problems may include, but are not limited to, slicing, droop, texture mapping issues and colour matching errors. All of these pose potential challenges when identifying facial expressions. Slicing can cut off relevant detail, droop caused by camera positioning can change visual appearance and can make one emotion appear like another, perhaps making the capture subject look sad [6]. Poor texture mapping impacts on visual quality and inadequate colour matching could cause a reconstruction to look over exposed or too dark. In order to compare 2D and 3D video media, a viewpoint was matched when recording from the 3D video stream. This allowed time stamped films to be created from each medium, to ensure that participants were given equal representations of non-verbal behaviour during the lie detection, allowing a fair comparison. The introduction of a free viewpoint at this stage would not have allowed this.

### **2.6 Research Question**

How do different avatar representations of the same motion data set impact upon trust?

## 2.7 Hypotheses

In an experiment a sender gave true and false accounts, each sent through both an avatar reconstructed from multiple videos, and through conventional video conferencing. Receivers then watched video footage of both. It was hypothesised that receivers would perform better using the footage of the video conferencing. The rationale is that the detail of the facial expression would be lost or distorted through the reconstruction process. This would make facial expressions harder to identify and, thus, make it more difficult to detect truthfulness.

### Hypotheses

When receivers observe footage of a 3D reconstructed avatar rather than video conferencing:

H1: The accuracy of determining truthfulness will be significantly less.

H2: Cognitive workload will be significantly higher when determining truthfulness.

While it is expected that the reconstructed avatar will be inferior to video conferencing, the experiment tests whether there will be sufficient visual quality to convey truthfulness.

## 3 Method

This experiment investigated whether avatars, reconstructed in real time using shape-from-silhouette, overcame the reported problem of people trusting CGI avatars too much. To test this, the gap between 3D reconstructed video and video conferencing in portraying deceit was measured. Two senders each gave a set of accounts of which half were true. These were captured and transmitted simultaneously across two mediums. Recordings of these sessions were later examined by sets of participants. As discussed in section 3.3 Detection of Deception,  $n=21$  receivers were asked to determine which accounts were true.

### 3.1 Scope

The scope of this experiment was to test if the impoverished facial representation of 3D reconstructed avatars, results in those using the medium to determine truthfulness, over trusting them. The reason that this might occur is that a lack of facial expressions has been shown to have this effect when using pre-modelled fixed expression authored abstract avatars or voice alone. To isolate the impact of visual reproduction from that of support for spatial qualities, the viewpoint has been fixed for the 3D video capture process, close to that seen through the 2D videoconferencing condition; thus not measuring the benefits of support for mutual gaze and interpersonal distance, instead leaving this for a later study. The study discusses the results together with a small number of previous studies to test what the researcher believes a survey of the literature suggests: that a lack of facial expression possibly increases truth bias. However, this was not

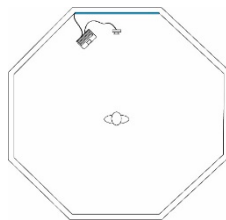
tested as rigorously as the researcher would have liked. The reason being that the experiment did not implement a third condition to directly test an authored abstract avatar. This was because, in order to record the facial tracking information needed, cameras would have had to be positioned closer to the face. These cameras would have occluded the capture subject from the 3D reconstruction cameras in the capture space. There are different methods available to capture and reproduce the face's impact on quality in different ways. It is likely that another approach, for example depth-based reconstruction, might produce different results. The results from this study may only be valid for shape-from-silhouette.

### 3.2 Session capture

In this section the set-up of the hardware configurations and session capture process is explained. Firstly, the environments that the sender and questioner occupied are described, in particular, how the information was captured and displayed in both, alongside a brief overview of the technology that linked these spaces through display and capture. Secondly, the telecommunication systems are described.

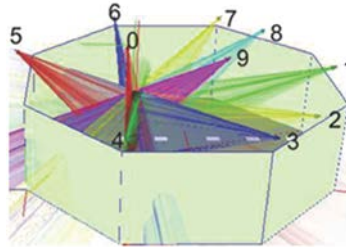
#### The sender environment

Each sender was captured within the OCTAVE display system [37] designed for telepresence research. The system uses ten HD cameras positioned at the top of the display walls. Figure 1 shows the capture space with someone being captured within it. Figure 2 diagrammatically illustrates the camera positions in the context of the walls and the capture subject, in addition it depicts the camera cones, from the silhouette of the capture subject from each camera image [38]. Only one screen of the immersive display used displayed the videoconference session; the others provided combined lighting and Chroma keying. The display used at the sender's end of the videoconference was a simulation grade active surface. Each of the eight display surfaces are 4x2m and use Christie S+3K Stereo DLP. They were run in mono with a resolution of 1400x1050 at 102Hz. During capture, defuse and ambient light from the OCTAVE's fourteen projectors (6 ceiling and 8 wall) and defuse spotlights were combined to achieve clear lighting and contrast of the capture subject's face. The room's strip lighting was also turned on.



**Fig. 1.** The sender environment. The sender stands in the centre of the capture suite. A Macbook Pro connected to a webcam captures their image and transmits it through Skype. Simultaneously, multiple cameras (**Fig. 2.**) are recording and streaming images of the sender for 3D reconstruction.

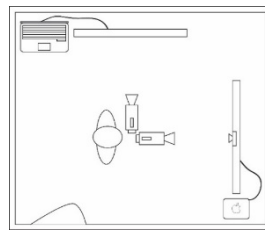




**Fig. 2.** Camera placement with respect to the sender position and display walls. Silhouette cones joins each camera to the silhouette edges of the sender as seen by that camera. Copyright IEEE [38].

### The environment shared by the questioner and session display and capture

The interviewer was in a separate office, along with two large televisions, each connected to a respective computer. One television showed the live video conference, the other a live 3D reconstruction. In both cases, the head and shoulders of the sender filled the screen. In one this embodiment was 2D video, in the other 3D a reconstructed avatar using shape-from-silhouette. A camera facing each screen recorded the respective embodiment and the sound coming from the videoconference link. A web cam was positioned above one of the televisions recording the questioner who stood in front of it. The video from the web cam was transmitted to the sender environment. Screen and camera update rates were both set to 50Hz, which allowed the two to work most effectively together.



**Fig. 3.** The environment shared by the questioner and the session display and capture. The 3D reconstruction was on the top screen and the video conferencing on the right.

### Telecommunication equipment and set up

The video conferencing and the 3D reconstruction system displayed the head and shoulders of the sender on identically large wide HD screens. Both systems captured the sender simultaneously. The audio from the video conferencing was used for both. There was no discernible difference in the latency of the two systems. Both gave good lip-sync.

The video conferencing software was Skype which, at the time of the experiment, was the prevalent commodity system, widely used at work and home. The version of Skype used was 6.7.0.373. The network was a wireless 1Gb private network under light load. The Skype was run at both ends on Mac Mini computers. These were connected via a 1Gb wireless network under low utilisation.

The 3D reconstruction system created a real-time avatar, utilising both geometry and texture, using shape-from-silhouette. The reconstruction method is outlined in [39]. The capture and pre-processing method is covered in [40] while the method of creating form is based on the EPVH algorithm [41]. The experiment utilised parallelisation to produce real-time frame rates, from 10 HD cameras, on one computer, as described in the study of [42]. The reconstructed avatar was formed, textured and viewed on it. Pairs of video streams were taken by five DELL dual core PCs. These ten streams were then sent to a MacBook Pro in the questioner's room. This laptop undertook image pre-processing, including generating silhouettes, reconstruction and rendering. It achieved this at between 20 to 30 fps dependent on the segmentation complexity of shape-from-silhouette from each camera.

## Procedure

Two senders were asked by a questioner to give true and false accounts to four questions. Each sender was asked to walk inside the capture system, stand close to an X on the floor, and face a display wall. Displayed on this wall was a questioner seen in an office through a videoconferencing session. The sender and questioner exchanged pleasantries. The questioner explained to each sender that they would be asked four questions, to which the sender should provide a true and false account. The sender was asked to make each account around one minute long, to randomise the order in which they told truth and lies, and not to tell the questioner which account was true until after all were given. The questioner then asked each question in turn, waiting between each response allowing two accounts to be given. The sender gave these accounts accordingly and, at the end, said which were true. The questions asked were:

Q1. "In under a minute each could you describe two accounts of your route to work today for me please?"

Q2. "In under a minute each could you describe two accounts of a sporting achievement from your past please?"

Q3. "In under a minute each could you describe two accounts of a comical story from your past please?"

One of the following two questions was specifically addressed to each sender in an attempt to add pressure:

Q4. "I have been sending you emails marked urgent and you have not yet responded, could you give me two accounts as to why?"

Q5. "You are in charge of the commercial partnership with body A. Could you give me two accounts as to how this relationship is going?"

Each account was transmitted simultaneously through the two mediums of 3D reconstructed video and video conferencing to two wide screen televisions in the questioner's room. Video from cameras facing these two screens was used to create 24 short clips.

### Subjects

Two senders were recruited. Each worked within the University in roles of commercial engagement. They were chosen as both had excellent communication skills. This was done in an effort to not bias the experiment through using a sender not competent at public speaking or an actor expert in deception through acting. The senders knew the author. One of them knew the questioner.



**Fig. 4.** Facial-expressions captured through Videoconferencing (VC) and Virtuality Telepresence (VT). Senders give true and false accounts, each simultaneously transmitted across the two mediums, video of which is watched by receivers asked to decide which are truthful.

### 3.3 Detection of Deception

This section outlines the design and implementation of the study to determine the ability of each medium to portray truthfulness.

#### Procedure

N=21 receivers watched the series of 24 videos, choosing which were true and false accounts. Each receiver entered a computer lab and sat at a desktop computer. A pad was placed on their forehead to measure cognitive load. A band holding it was tightened around their head. Wires were attached. A head scarf was wrapped over this equipment to shield it from light and perhaps make the participant less self-conscious. They were asked to follow the instructions on the screen. These told them to watch a series of videos, selecting after each either a button for true or false, based on their overall judgment of the account. They were also asked to press a button each time they thought they detected a lie; however, this data was not used as many reported later to have delayed their decision.

## Subjects

The receivers were recruited from University staff and students: 19 were male and 2 were female. The number of male and female subjects was not matched as they were not interacting with the senders. All were healthy and had the mean age of 37.38 with standard deviation of  $\pm 9.73$  years. Recruitment was undertaken via email. Most worked or studied at the University. 17 were staff and 4 were students. The experiment lasted approximately 20 minutes. The participants were not paid for taking part.

## Receiver environment

Each receiver sat in the same computer lab but at a different time. The videos were displayed using E-Prime [43] for randomisation and data collection on a standard desktop computer with inbuilt speakers for audio. A continuous wave functional near infrared spectroscopy (fNIRS) system [44] was utilised to measure prefrontal cortex activity. fNIRS is similar to functional magnetic resonance imaging in that it provides a hemodynamic measurement of brain activity. fNIRS employs optical signalling to record oxygenated and deoxygenated haemoglobin concentration changes which, in turn, can be used to provide a measure of cognitive workload. By recording optical density changes at two differing wavelengths (700nm and 830nm) the relative change over time in oxygenated and deoxygenated haemoglobin can be calculated using the modified Beer-Lambert Law.

Data was collected at 2Hz using Cognitive Optical Brain Imaging (COBI) Studio software [45]. Raw fNIRS light intensities were subjected to signal processing to remove any motion or physiological artefacts (heart rate or respiration rate) by employing a linear phase low pass finite impulse response filter with a cut off frequency of 0.1Hz.



**Fig. 5.** Participant environment. The participant is sat facing a desktop machine that is running the sender movies in E Prime. The participant is wearing an fNIRs sensor with a scarf wrapped around it to ensure light did not affect the data captured.

## 3.4 Variables

The dependent and independent variables are described below.

### Independent Variables

The conditions looked at were medium and truthfulness. The medium of video conferencing was tested against that of virtuality telepresence. Two senders were each asked to give both a true and false account in answer to four questions.

**Table 1.** Independent variables

	Videoconferencing	Virtuality Telepresence
Truthful account	VCt	VTt
Deceit- ful account	VCd	VTd

### Dependent Variables

The dependent variables were the success of determining truthfulness in each medium and changes in the level of cognitive effort while doing so. The scoring scheme was simple in that the correct identification of truthfulness (i.e. the correct identification of a truthful or a deceitful account) was scored with 1 and an incorrect judgement with 0. Thus, for a given medium, given that there were 6 accounts, if all truths and lies were correctly identified a maximum score of 6 would be obtained, and if none were identified the participant would obtain a total score of 0. In addition, data on cognitive load using fNIRS were collected while the receivers conducted the task. This data measured the change in the oxygenation of the blood in the prefrontal cortex. Increases in oxygenation are an indication of an increase in cognitive load.

**Table 2.** Dependent Variables

	Description	Range	Type
Score	Number of successes in determining the truthfulness of account	0 to 6	Task performance
Oxygenation	Change in oxygenation of the blood in the prefrontal cortex	-1 to 5	Difficulty

### 3.5 Ethics

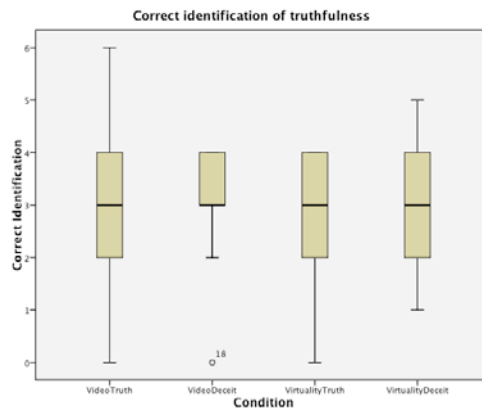
Ethical approval was gained for the study from the University ethics board. The approval number was CST 12/51. The ethical approval included approval of the information sheet and the consent form. The information sheet explained to the participants the general theme of the study and what they might expect. The consent form informed

participants of the anonymity of data, the right to leave at any time and the right to have one's data completely removed.

## 4 Results

A quantitative analysis looked at task performance and an indication of cognitive load, and then correlated the two.

When testing H1 (Accuracy of determining truthfulness will be significantly less when determining truthfulness through the reconstructed avatar), no significant difference was found. This hypothesis was tested using a paired samples t-test and rejected ( $t= 1.072, p=0.297, df=20$ ), see Graph 3 below.

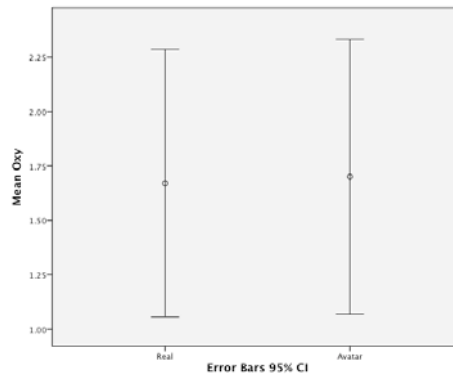


**Graph 3.** Task performance - correct identifications of “truthfulness”

**Table 3.** A table showing the correct identification of truth and deceit. The totals are out of a possible 126 statements.

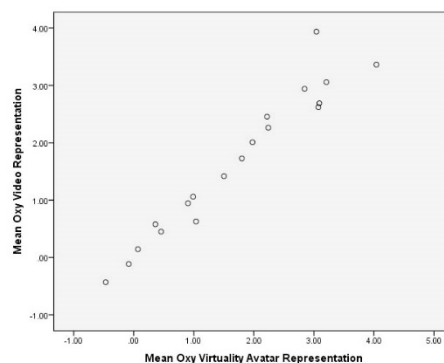
Real		Reconstruction	
Truth	Deceit	Truth	Deceit
66	64	58	63

When testing H2 (Cognitive workload will be significantly higher when determining truthfulness through the reconstructed avatar), the result was that cognitive load, as indicated by change in the oxygenation of blood in the frontal cortex, under the two conditions, was not found to be significant ( $t= -0.40, df=18, p=0.694$ ). See Graph 4 below.



**Graph 4.** Comparing change in cognitive load across the two mediums - mean change in the oxygenation of blood in the prefrontal cortex for two groups

An interesting follow up to this finding is that when the two variables are correlated using the Pearson product-moment correlation coefficient the finding is highly significant ( $r=0.968$ ,  $N=19$ ,  $p<0.01$ ). This finding strongly suggests that receivers who found the task difficult when looking at the video had a similar difficulty when looking at the virtuality telepresence version, and vice versa. This would indicate that the virtuality telepresence version is a very good virtual equivalent of the video version.



**Graph 5.** Mean change in oxygenation when viewing video representation against mean change in oxygenation when viewing virtuality avatar representation during the detection of truths.

## 5 Discussion

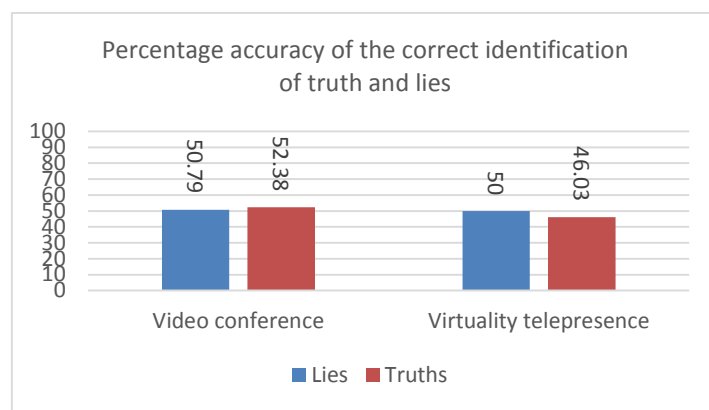
This experiment set out to determine whether 3D reconstructed avatars created using shape-from-silhouette displayed the same truth bias that authored abstract 3D avatars were shown to in a comparison of a small number of previous studies. The researcher suggests that talking to a user wearing an abstract 3D avatar, with no facial animation, is similar to talking to somebody wearing a mask. Therefore the research questions were firstly: Would a 3D medium that removed this mask, result in a truth bias akin to

video and thus the real world? Secondly, do the shortcomings of the tested method for reconstruction contribute to a distortion in truth bias? In order to test these questions the virtuality telepresence medium first needed to be tested independently. This was achieved by comparing the virtuality telepresence condition against that of video conferencing, testing whether the visual quality of the medium would impact upon a receiver's ability to detect truthfulness. The results from this experiment would then allow a comparison of the medium with those tested in previous studies from both VR and psychology. The following hypotheses were tested and the outcome for each hypothesis is presented.

H1: The accuracy of determining truthfulness will be significantly less when viewing the reconstructed avatar – Overall there was no significant difference in accuracy between video conferencing and virtuality telepresence.

H2: Cognitive workload will be significantly higher when determining truthfulness when viewing the reconstructed avatar - There was found to be no significant difference in levels of cognitive effort when identifying truthfulness across media. There was, however, a significant correlation between workload across each medium, suggesting that receivers have to work equally hard when viewing either medium.

No significant difference was found in the accuracy of determining truthfulness across the conditions ( $t= 1.072$ ,  $p=0.297$ ,  $df=20$ ). This was the case when the sender was telling the truth or lies, and overall. There was no significant difference in the number of accounts considered to be true or false ( $t= 1.072$ ,  $p=0.297$ ,  $df=20$ ). This suggests that there was no difference in truth bias across the conditions. These results are consistent with other studies using videos but different to those using CGI avatars (with limited facial animation) or voice alone. However, a negative truth bias was evident between the 3D video based medium and previous psychological studies. The significance of the latter was not testable, due to the detail of reporting in the literature. There was a significant correlation in changes in cognitive workload between conditions ( $r=0.968$ ,  $N=19$ ,  $p<0.01$ ). This suggests that individuals worked equally hard to determine the truthfulness from the footage of both mediums. This suggests that the 3D medium did display a sufficient quality of representation for observers to determine the truthfulness at a similar level as that for video conferencing.



**Graph 6.** Accuracy of determining truthfulness across the media



*Graph 6.* visualises the accuracy of determining truthfulness, for the true statements and the false statements. Both mediums have an average overall accuracy close to chance (50%). Video conferencing showed a combined average of 51.58% and virtuality telepresence 48%. These are close to the 54% average of the real world studies reported by Bond and DePaulo [1]. The reported differences in accuracy when identifying truth and lies is also small. It is much smaller than those shown in the avatar study reported by Steptoe et al. [2] which lacked facial expressions. The avatar study by McDonnell et al. [3] that did reproduce facial expressions, reported a truth bias close to real world studies. It should be noted that, due to the small number of studies available for comparison and the relatively small number of participants when compared to real world studies, a much larger study testing all conditions would be needed to draw any firm conclusions. The above results were promising for the medium of virtuality telepresence. They show that, even with the impacting factors of shape-from-silhouette, virtuality avatars contain enough detail to allow the receiver to ascertain truthfulness comparably with video conferencing.

## 6 Summary

This experiment set out to compare the detection of deceit across 2D and 3D mediums which were designed to faithfully represent appearance and identity. The viewpoint of the latter was fixed to isolate the impact of the quality of facial representation from mutual gaze. A free viewpoint would have allowed the user to view the avatar from any angle making a direct comparison of quality impossible. In addition, it would have added an interaction task that would have impacted on the cognitive workload of the participant. The first contribution is to demonstrate, through an experiment, that a 3D medium aimed at faithfully reproducing a person's appearance can be comparable to video conferencing in the overall portrayal of deceit. The second contribution is to demonstrate through an experiment that participants work equally hard when detecting deceit through a 3D reconstructed medium and video conferencing. The third contribution is to link the two to provide additional evidence to support the above trend.

In addition, there are three methodological contributions:

1. Locking and matching a viewpoint to allow exact contrast of visual quality without the distraction of interaction.
2. An experimental setup for simultaneously capturing 2D video and video based 3D reconstructed video for contrast purposes.
3. The use of fNIRS to compare cognitive effort during the detection of truths in 2D video and video based 3D reconstruction for contrast purposes.

This experiment has shown that a 3D graphics avatar, continually reconstructed from live video, is comparable to video conferencing in portraying the overall truthfulness of the person it mimics. A comparison with the small amount of previous literature on this subject suggests that people trust others too much when a medium does not reproduce their facial expression; however, a much larger study would be needed to draw any firm conclusions. In addition, it would be interesting to examine the impact of a

free viewpoint and to explore how the ability to move around the sender would impact upon how well the receiver is able to detect deception and identify body language, thus testing whether freedom of movement and ability to interact is a help or a hindrance.

These findings may contribute to the design of future virtual humans and communication mediums. They can also be relevant regarding the choice of mediums used in the justice system, in advertising, social media, politics, distributed family and the workplace. Not everyone will want to remove the mask but perhaps further research can inform others of its influence.

## 7 References

1. CF Jr Bond, B.D.: Accuracy of deception judgments. *Pers Soc Psychol Rev* 10, 214-234 (2006)
2. Steptoe, W., Steed, A., Rovira, A., Rae, J.: Lie tracking: social presence, truth and deception in avatar-mediated telecommunication. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1039-1048. ACM, (Year)
3. McDonnell, R., Breidt, M., Bulthoff, H.: Render me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. In: *ACM Transactions on Graphics*, pp. 1-11. (Year)
4. Bailenson, J.N., Beall, A.C.: Transformed social interaction: Exploring the digital plasticity of avatars. *Avatars at Work and Play*, pp. 1-16. Springer (2006)
5. Ekman, P.: *Telling lies: Clues to deceit in the marketplace, marriage, and politics*. New York: Norton (1985)
6. Roberts, D.J., Fairchild, A.J., Champion, S.P., O'Hare, J., Moore, C.M., Aspin, R., Duckworth, T., Gaspardo, P., Tecchia, F.: withyou—An Experimental End-to-End Telepresence System Using Video-Based Reconstruction. *Selected Topics in Signal Processing, IEEE Journal of* 9, 562-574 (2015)
7. Vrij, A., Mann, S.: Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology, Crime and Law* 7, 119-132 (2001)
8. Horn, D.B.: Is seeing believing?: detecting deception in technologically mediated communication. In: *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, pp. 297-298. ACM, (Year)
9. Frank, M.G., Feeley, T.H.: To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research* 31, 58-75 (2003)
10. McKenzie, F.R., Scerbo, M., Catanzaro, J., Phillips, M.: Nonverbal indicators of malicious intent: affective components for interrogative virtual reality training. *International journal of human-computer studies* 59, 237-244 (2003)
11. Ekman, P., Friesen, W.V.: Nonverbal leakage and clues to deception. *Psychiatry* 32, 88-106 (1969)
12. Ekman, P.: Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences* 1000, 205-221 (2003)
13. Feldman, R.S., Jenkins, L., Popoola, O.: Detection of deception in adults and children via facial expressions. *Child development* 350-355 (1979)
14. Millar, M.G., Millar, K.U.: The effects of cognitive capacity and suspicion on truth bias. *Communication Research* 24, 556-570 (1997)

15. Ekman, P.: Facial expression and emotion. *American psychologist* 48, 384 (1993)
16. Endres, J., Laidlaw, A.: Micro-expression recognition training in medical students: a pilot study. *BMC medical education* 9, 47 (2009)
17. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 3549-3557 (2009)
18. Porter, S., Ten Brinke, L.: Reading between the lies identifying concealed and falsified emotions in universal facial expressions. *Psychological Science* 19, 508-514 (2008)
19. Bernstein, D.M., Loftus, E.F.: How to tell if a particular memory is true or false. *Perspectives on Psychological Science* 4, 370-374 (2009)
20. Ekman, P., Friesen, W.V.: *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press (1978)
21. Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R.C., Mathiak, K.: Recognition profile of emotions in natural and virtual faces. *PLoS One* 3, e3628 (2008)
22. Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B.: Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36-47 (2012)
23. Ferrari, M., Quaresima, V.: A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63, 921-935 (2012)
24. Hu, F., Wang, Y., Cao, X., Hao, Q., Brown, D.: Using Wiimote and Kinect for Cognitive Rehabilitation: Toward an Intelligent Sensor/Image Signal Processing. *Cyber-Physical Systems: Integrated Computing and Engineering Design* 321 (2013)
25. Seraglia, B., Gamberini, L., Priftis, K., Scatturin, P., Martinelli, M., Cutini, S.: An exploratory fNIRS study with immersive virtual reality: a new method for technical implementation. *Frontiers in human neuroscience* 5, (2011)
26. Cansiz, Y., Tokel, S.T.: Effects of Way Finding Affordances on Usability of Virtual World Environments in Terms of Users' Satisfaction, Performance, and Mental Workload: Examination by Eye-Tracking and fNIR. In: *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp. 1073-1079. (Year)
27. <http://www.vicon.com/Press/Details/10>
28. <http://www.faceshift.com/>
29. Fuchs, H., Bishop, G., Arthur, K., McMillan, L., Bajcsy, R., Lee, S., Farid, H., Kanade, T.: Virtual space teleconferencing using a sea of cameras. In: *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*. (Year)
30. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., Fuchs, H.: The office of the future: A unified approach to image-based modeling and spatially immersive displays. In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 179-188. ACM, (Year)
31. Roberts, D., Rae, J., Duckworth, T., Moore, C., Aspin, R.: Estimating the Gaze of a Virtuality Human. *IEEE Transactions on Visualization and Computer Graphics* 19, 681-690 (2013)
32. Maimone, A., Fuchs, H.: Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In: *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pp. 137-146. IEEE, (Year)
33. <http://www.microsoft.com/en-us/kinectforwindows/>

34. Pan, X., Gillies, M., Slater, M.: The impact of avatar blushing on the duration of interaction between a real and virtual person. In: Presence 2008: The 11th Annual International Workshop on Presence, pp. 100-106. Citeseer, (Year)
35. Spellman, B.A., Tenney, E.R.: Credible testimony in and out of court. *Psychonomic bulletin & review* 17, 168-173 (2010)
36. Bailenson, J.N., Blascovich, J., Beall, A.C., Noveck, B.: Courtroom applications of virtual environments, immersive virtual environments, and collaborative virtual environments. *Law & Policy* 28, 249-270 (2006)
37. <http://www.salford.ac.uk/computing-science-engineering/facilities/octave-technical-information>
38. Roberts, D.J., Rae, J., Duckworth, T.W., Moore, C.M., Aspin, R.: Estimating the gaze of a virtuality human. *Visualization and Computer Graphics, IEEE Transactions on* 19, 681-690 (2013)
39. Duckworth, T., Roberts, D.J.: Parallel processing for real-time 3D reconstruction from video streams. *Journal of Real-Time Image Processing* 9, 427-445 (2014)
40. Moore, C., Duckworth, T., Aspin, R., Roberts, D.: Synchronization of images from multiple cameras to reconstruct a moving human. In: *Distributed Simulation and Real Time Applications (DS-RT), 2010 IEEE/ACM 14th International Symposium on*, pp. 53-60. IEEE, (Year)
41. Franco, J., Boyer, E.: Efficient polyhedral modeling from silhouettes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31, 414-427 (2009)
42. Duckworth, T., Roberts, D.J.: 3DRecon A utility for 3D reconstruction from video. short paper at the Joint Virtual Reality Conference (ICAT, EEG, Eurographics), Madrid (2012)
43. <http://www.pstnet.com/eprime.cfm>
44. <http://fnirdevices.com/>
45. <https://www.biopac.com/Manuals/fNIR%20COBI%20Manual.pdf>