Virtual Blindness - A Choice Blindness Experiment with a Virtual Experimenter

Martin Lingonblad[1], Ludvig Londos[1], Arvid Nilsson[2], Emil Boman[2], Jens Nirme[1], and Magnus Haake[1(✉)]

[1] Division of Cognitive Science, Lund University, Lund, Sweden
{martin.lingonblad.299, fte08llo}@student.lu.se,
{jens.nirme, magnus.haake}@lucs.lu.se
[2] Department of Computer Science, Lund University, Lund, Sweden
{ada09an2, emil.boman.633}@student.lu.se

Abstract. How are people facing a virtual agent affected by the vividness and graphical fidelity of the agent and its environment? A choice blindness (CB) experiment – measuring detection rate of hidden manipulations – was conducted presenting a high versus low immersion virtual environment. The hypothesis was that the lower quality virtual environment (low immersion) would increase the detection rate for the CB manipulations. 38 subjects participated in the experiment and were randomized into two groups (high and low immersion). Both conditions presented a virtual agent conducting the CB experiment. During the experiment, 16 pairs of portraits were shown two at a time for the participants who were then asked to choose which portrait they found most attractive. For eight of the pairs, participants were asked to justify their choice while in four cases their choice had been secretly switched to the portrait they had not chosen. If a participant stated that the chosen portrait had been switched, it was annotated as a concurrent detection.

The results revealed an increase in detection and earlier detection rate for the low immersion implementation compared to the high immersion implementation. Future research may involve experiments with higher degree of both immersion and presence, using for example head mounted display systems.

Keywords: Virtual agent · Choice blindness · Attention · Presence · Immersion

1 Introduction

1.1 Virtual Environments

A virtual environment (VE) provides control over a setting, similarly to a laboratory environment, with the advantage that a VE might emulate a more natural setting [1].

VEs also have the advantage that virtual humans or virtual agents can be used in the role as experimenters. The virtual agents will act the same way for each participant, something that is hard to control for with real humans as experimenters [2].

However, the validity of experiments in virtual environments in many cases remains unconfirmed. Do we act similarly in a VE as we would in the real world?

W.-P. Brinkman et al. (Eds.): IVA 2015, LNAI 9238, pp. 442–451, 2015.
DOI: 10.1007/978-3-319-21996-7_47

And more specifically; do we interact with a virtual human similarly to how we interact with other humans?

1.2 Immersion and Presence

In virtual reality the terms *immersion* and *presence* [3, 4] are often used to determine the degree of match to reality. *Presence* refers to a subjective measure of a person's experience of being in a virtual environment and can be subdivided into environmental presence, co-presence, and social presence [3].

There is evidence suggesting that participants who experience a high level of presence to a larger extent also behave like they would in a natural setting, including interaction with (virtual) humans [3]. A well-animated agent can be perceived as "human-like, engaging, and motivating" [2]. Furthermore, requirements for realistic graphics and animation are higher for interacting with a virtual agent as if it was a human, compared to interacting with an avatar (a virtual representation of another human [1, 2]. *Immersion* comprises the technical parts of the virtual setting. Slater and Wilbur divide immersion into four categories: *vivid*, *inclusive*, *extensive*, and *surrounding* [3]. "Vivid" refers to technical specifications such as resolution and fidelity. "Inclusive" refers to how much of reality is shut out. "Extensive" refers to how many sensory modalities are included. "Surrounding" refers to how well the visual field is represented. *Immersion* and *presence* are casually related. Welch et al. reported that high or low graphical fidelity affected participants' experienced presence during a car simulator experiment [5]. Likewise, Shubert et al. concluded: "The more inclusive, extensive, surrounding, and vivid the VE is […] or the more similar the transformations in the VE are to those in the real world […] the higher the presence." [6].

High immersion, however, does not necessarily lead to high presence. Another important factor is the ability to convince oneself that the virtual reality is real – something commonly referred to as *suspension of disbelief* [7]. An example of low immersion but high presence is reading an engaging book.

In the current study we exploited the Choice Blindness research paradigm [8, 9] where aspects of the physical environments are manipulated by real human experimenters using techniques inspired by the domain of close-up card-magic. A central measurable parameter is participants' detection-rate of the manipulations. The current study addressed whether the level of immersion would affect perception of and interaction with a virtual agent experimenter, as reflected in the detection rate of manipulations.

1.3 The Choice Blindness Paradigm

In the original version of a CB-experiment [8] a participant is shown a number of different photographs of human faces displayed in pairs, during 2 to 5 s intervals. After the view period the photographs are turned face-down on the table in front of the participant and the participant is asked to indicate the photograph he or she preferred. Then the experimenter pushes the chosen card towards the participant who picks it up and looks at it and is asked to justify her choice. In some of the trials, the experimenter has swapped the faces by having hidden copies of both photographs in each hand. In these manipulated trials, the participant is forced to justify a choice that has not been made. For 70–80 % of the trials with manipulation go undetected.

In studies where the original experimental set-up with a real experimenter and real cards, was changed into a real experimenter using digital cards displayed on a screen [10] and into a virtual experimenter using digital cards [11] detection rates increased, with 60–65 % of the manipulations undetected.

Why have participants more often noticed the manipulation when the photographs have been displayed digitally on a screen instead of physically on a table? One possibility is that the natural environment is more trusted: real photographs usually are not suddenly changed into another. Another possibility is that the natural environment is more distracting than the virtual and more likely to disrupt participants' attention.

In the current study, we will utilize CB to see if participants' detection rates are affected by the level of immersion in a virtual environment. The study is a conceptual replication of the original Choice Blindness experiment [8], but instead of exploring the choice blindness phenomenon itself – it uses the choice blindness paradigm as a measurement of the experience when interacting with a virtual agent. We hypothesize that the level of immersion will affect detection rate and possibly shade light on the roles of (i) trust in a real versus virtual environment and (ii) distractibility of a real versus virtual environment in affecting detection rates.

2 Method

2.1 Design

Participants were assigned to one of two virtual implementations of the *Choice Blindness* paradigm [8, 9]. One group was presented with a low immersion implementation and the other group with a high immersion implementation. Before the virtual choice blindness task, the participants were tested for individual tendencies to experience presence via a translated and adapted version of the *Immersive Tendencies Questionnaire* [12]. After the virtual choice blindness task, participants were debriefed via a standardized choice blindness debriefing format.

2.2 Participants

41 university students (20 male, 21 female), age 19 to 32 years, were recruited at Lund University, Sweden. The majority were students at the Faculty of Technology.

3 participants were excluded from the data set due to not accurately completing the task or a bug in the program that affected their performance. The final data set included data from a total of 38 participants (21 females and 17 males) with 19 participants in each condition, cf. Table 1.

Table 1. Participants: number, gender, age (ranges, medians and averages).

Condition	N	N (female /male)	Min – Max (age)	Median (age)	M (age)
Low immersion	19	11 /8	19 - 31	24	23.8
High immersion	19	10 /9	20-32	24	24.2

2.3 Procedure

After having read and signed a written consent form the participants were presented with a questionnaire addressing individual tendencies to experience presence. Next, the participants were told that they were to perform a task interacting with a virtual agent using a voice recognition system. This instruction was delivered to conceal the actual experimental design. The participants were then escorted into a booth where a computer with the virtual choice blindness task was prepared (having a "Wizard-of-Oz" [13] hiding in an adjacent booth).

Participants were assigned to one of the two conditions depending on the result of the ITQ together with a randomization table. The ITQ-result was used to ensure matched groups in the two conditions with regard to immersive tendency and to control for computer experience.

The virtual choice blindness task consisted of 16 sessions. In each session the virtual experimenter presented a pair of photographs showing portraits, with four of these sessions being manipulated according to the choice blindness paradigm [8, 9].

After completion the participants were brought back to the entering room and debriefed using a standardized choice blindness debriefing procedure [8, 9]. The experiment was finally concluded by giving the participants a restaurant voucher.

2.4 Immersive Tendencies Questionnaire

The widely used *Immersive Tendencies Questionnaire* (ITQ) [12] is designed to measure the tendency of individuals to be involved or immersed in a virtual environment by addressing several factors that are assumed to contribute to immersive tendencies, among others the tendency to become involved in activities, the tendency to maintain focus on current activities, and the tendency to play video games.

ITQ consist of 18 items, each with a seven-point scale semantic differential scale complemented with a midpoint anchor. In this study, the questionnaire was translated to Swedish with an additional 19th item probing for computer experience.

2.5 Virtual Setting

Design of the Virtual Scenario: The CB paradigm was implemented in a VE with a virtual experimenter and digital photos (from the original choice blindness card collection). As in the original studies, each photo pair was presented during 3 s. The size of the virtual cards was adjusted to resemble the size experienced in the original (real world) experiments when "virtually" presented on the screen.

Following the choice blindness protocol, the participant was asked to verbally describe their choice on 8 of 16 pairs during the experiment. Of these 8 pairs, 4 were manipulated so that the photo of choice was secretly (digitally) switched whereupon the participant actually was asked to describe why they preferred the photo they did not chose.

Main Measurement: The core measure within the choice blindness paradigm is *concurrent detection*. In case of explicit verbal reactions from the participant, a concurrent detection can be clearly registered. To handle more subtle cases, participants may, as in the present study, be video recorded during the experiment for post-experiment discern whether a detection of the manipulation was made. In addition there is a standardized debriefing procedure after the choice blindness experiment. During this debriefing, the participants are given a chance to make a post-experiment detections guided by a specified question procedure where the questions are asked step by step – revealing more and more information about the experiment.

Realization of the Virtual Scenario: The virtual implementation was developed based on motion capture recordings with one of the experimenters from the original choice blindness experiments going through the experimental procedure. An eight camera Qualisys® motion capture system was used; the recordings were imported into Autodesk MotionBuilder® and plotted onto a 3D-character created in Autodesk Character Generator®. The motion capture recordings were then edited into shorter clips representing different animated sequences of the experimental procedure, for example welcoming the participant, showing the pair of cards, asking the participant to describe a certain choice, etc. Next, the animated clips were imported into the Unity® developer platform together with 3D-models of the furnishings created in Autodesk 3ds Max®, whereupon the final implementation of the virtual scenarios were edited and scripted. As a real interactive implementation of a virtual agent with a spoken natural language interface was beyond the scope of this experiment, the actual interaction was simulated using a “Wizard-of-Oz”-approach [13]. In this realization, the participant communicated with the virtual experimenter through a head-set while the “Wizard” was hidden in an adjacent booth listening to the participant while triggering the verbal utterances and movements of the virtual presenter. As for the two implementations of the virtual experiment (high versus low immersion) the aim of the high immersion setting was to replicate one of the original experimental environments and procedures as closely as possible, while the low immersion setting was a tuned down version of the high immersion setting. The goal with the low immersion version was to reduce experienced presence.

In the low immersion version (cf. Fig. 1), all superfluous furnishing of the virtual environment were removed according to the involvement parameter, thus lowering inclusiveness [3]. Furthermore, for the low immersion version, the voice of the virtual experimenter was “computerized” (tuned to sound like a computer generated synthetic voice) and the facial animations were simplified in line with the spatial presence parameter in order to reduce extensiveness [3]. Finally, in the low immersion version, shadows were removed [14] and the graphical settings were decreased in compliance with the perceived realism parameter to decrease vividness [3].

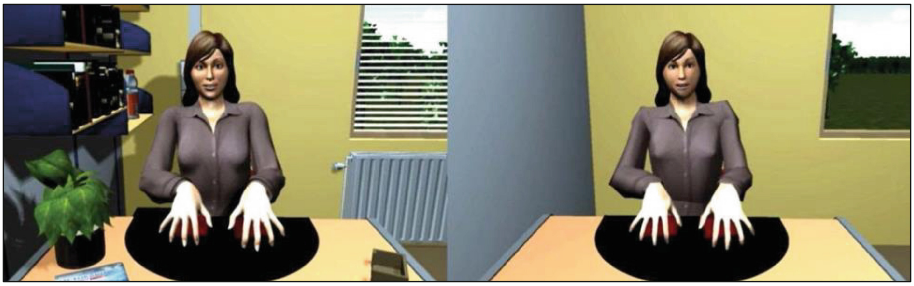

Fig. 1. Screenshots of the virtual implementation of the choice blindness experiment. Left: high immersion; right: low immersion.

A pilot study was conducted in order to validate that the two immersion conditions actually produced a difference in perceived presence and to reassure that the alterations in immersion did not affect the execution of the choice blindness task. The pilot study was performed in four steps, involving 10 students recruited from the same area (a campus restaurant) as the following experiment.

First, the participants were asked to fill out the *Immersive Tendencies Questionnaire* (ITQ) [12] to determine to what extent the participants tended to involve themselves in a virtual world. Second, they were randomly exposed to one of two virtual implementations of the choice blindness task (low or high immersion). Third, after the interaction with the virtual implementation, the participants were asked to fill out a questionnaire to evaluate their perceived presence during the intervention. The presence questionnaire was based on the *The Igroup Presence Questionnaire* (IPQ) [6] and is made up of questions addressing three aspects of presence (*spatial presence*, *involvement*, and *perceived realism*); five additional questions were added to the presence questionnaire to include *social presence* [15]. The fourth and final step was a standardized choice blindness debriefing procedure for evaluation of concurrent detection [8].

The ITQ-results showed an average of 0.62 in the high immersion condition and 0.65 in the low, suggesting that the two groups of participants were equal with regard to individual tendencies to involve themselves in virtual environment settings.

The results of the presence questionnaire (cf. Table 2) indicated a difference between the two implementations in the expected direction, with higher perceived presence for all four parameters compared to the low immersion implementation.

The evaluation of concurrent detection of the choice blindness data also showed a lower rate of detections in the high immersion implementation compared to the low immersion implementation.

Table 2. Average values of the presence questionnaire parameters for the two implementations of immersion.

Group	Spatial	Involvement	Exp. realism	Social presence
High immersion	2.28	2.05	1.65	2.52
Low immersion	1.84	1.30	1.15	2.16

Neither version showed any tendencies to hinder or affect participants negatively while they performed the choice blindness task.

2.6 Statistical Analyses

All analyses were conducted in R v3.1.3 [16]. Independent student's t-tests were used for overall comparisons of concurrent detection between immersion conditions. Multilevel logistic regression (R package: *lme4*) was used to compare the two immersion conditions with respect to the order of the manipulation at which the concurrent detections were registered. The alpha level for all statistical analyses was set to .05.

3 Results

3.1 Concurrent Detection

Concurrent detection [8, 9] was individually coded by two blinded persons from the research team. Following the coding, the results were compared and discrepancies were discussed and agreed upon while still blinded.

Concurrent detection was coded in two levels, *hard* and *weak* detection [8, 9]. Hard detection refers to any verbal detection ranging from stating clearly that the photo was not the one chosen to a vaguer indication, for example "Was this really the photo I chose?" Weak detection refers to a participant either frowning, looking puzzled, or behaving differently compared to non-manipulated trials. Weaker detection is not as evident as the harder detection, as it could be the case that the participant is not really aware of the manipulation. It could also be the case that the participant was aware but did not react verbally (cf. Fig. 2).

The total result for hard concurrent detection was for high immersion 27 % (including weak detection = 31 %), and for low immersion 38 % (including weak detection = 43 %). A straightforward t-test showed no significant results between the low and high immersion condition (hard detection: $t(150) = 1.483$, 95 % CI [-0.07; 0.5], $p = 0.14$).

A look at Fig. 2 suggests, however, a seemingly evident difference between the two groups (low and high immersion) over the course of the four manipulations. While concurrent detections were more or less equal at the first manipulation, the low immersion group encountered more frequent concurrent detections during the second and third manipulation. During, the fourth and last manipulation, the high immersion group "caught up" and even surpassed the low immersion group. A multilevel logistic regression analysis of the detections with regard to the order of the four manipulations revealed a significant difference between the low and high immersion group ($Z = 2.363$, $p = .018$) in that the low immersion group had both more and earlier concurrent detections compared to the high immersion group. Additionally, the multilevel logistic regression also indicated a significant interaction effect between the type of immersion and the order of the four manipulations ($Z = -2.149$, $p = 0.032$) supporting the finding that the low immersion condition had substantially earlier detections than the high immersion condition (cf. Fig. 2).

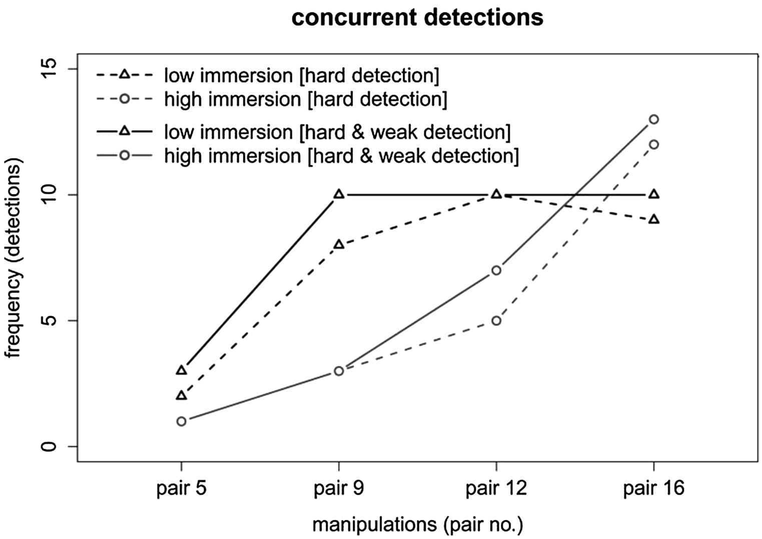

Fig. 2. Concurrent detection for low (green) and high (red) immersion. CD1 to CD4 (Current Detection 1 to 4) represents the four manipulations in the choice blindness experiment.

ITQ, the participant's extent of involving him/herself in a virtual world, were used to determine that there was no a misallocation between the high and low immersion groups. The mean ITQ for high immersion was 0.63 and for low immersion 0.59. A *t*-test showed no statistical difference between the groups.

4 Discussion

The study did not show any statistical difference in total detection rate between high and low immersion. However, the results indicated another difference in detection between the two conditions, and it is possible or even likely that a larger-scale experiment with more participants in each category would support our original hypothesis that low immersion will lead to a higher degree of detections in the CB task compared to high immersion.

As can be seen in Fig. 2, the patterns for *when* a detection is made differs between the two conditions. In the low immersion group detection in general came earlier than in the high immersion group. A possible reason for this result is that there were more interesting stimuli in the high immersion condition. In the low immersion condition there was a lack of stimuli to pay attention to, which could be the reason for the earlier detection. In other words, there was not many interesting things besides the faces to focus at and thus relatively more attention, compared to in the high immersion condition, was directed to the task at hand.

The concurrent detection results seem to be in par with similar research in the CB paradigm. For instance, the original CB study [8] reports a total detection rate at 26 %. For the high immersion condition in the present experiment the rate was 27–31 %, depending on how strict you want to be in calculating detection. We believe this shows that it is possible to recreate the CB effect in a virtual setting without producing a negative effect on the task. However, one should be vary of the design of the virtual setting. Our study does show an increase in detection and earlier detection rate for a low immersion design compared to a high immersion design. This may indicate that if one wants to replicate or design experiments in a virtual environment aiming at having participants behave naturally, one should take the time necessary to create a virtual setting which is as similar as possible to a real world setting.

5 Future Research

We would like to replicate the experiment with more participants and compare our results, foremost in the high immersion condition, with the earlier mentioned experiments by [10]. Another interesting possibility would be to repeat the study with a much more pronounced drop in graphical fidelity for the low immersion environment.

Using an intelligent virtual experimenter instead of a WOZ-setup with a human controlling the agent is also on our agenda. This would bring more control and ensure equivalence of vocal responses for all participants.

The experiment platform that we created can also be used for further and extended investigations. One may for instance add more properties to increase the immersion with the intention to increase the perceived presence. Possibly this might lower detection rate even more. However, that begs the question whether one could lower detection rate more. In one imagined scenario an *Occulus Rift* would be incorporated into the program to increase immersion. Would this lower the detection rate? We believe that the effect may be the opposite in making participants more suspicious and even increase detection. Imagine yourself being asked to do a “simple” face preference task but also be told that in order to do this simple task you need to use an advanced head mounted display. Yet we still find that it would be an interesting branch that this project could investigate.

It would also be possible to reduce the immersion level even more either by reducing all aspects or by focusing on one or some aspects, allowing distinguishing the most important aspect(s) behind the observed effect. In theory we could lower just the social presence aspect and see if that alone affects the detection rate. But our suspicion is that it is a combination of aspects that creates the CB effect.

References

1. Blascovich, J., Loomis, J., Beall, A.C., Swinth, K.R., Hoyt, C.L., Bailenson, J.N.: Immersive virtual environment technology as a methodological tool for social psychology. Psychol. Inq. **13**(2), 103–124 (2002)
2. Dehn, D.M., Van Mulken, S.: The impact of animated interface agents: a review of empirical research. Int. J. Hum Comput Stud. **52**(1), 1–22 (2000)

3. Slater, M., Wilbur, S.: A framework for immersive virtual environments (five): speculations on the role of presence in virtual environments. Presence Teleoperators Virtual Environ. **6** (6), 603–616 (1997)
4. Slater, M.: Measuring presence: a response to the witmer and singer presence questionnaire. Presence Teleoperators Virtual Environ. **8**(5), 560–565 (1999)
5. Welsh, R.B., Blackman, T.T., Liu, A., Mellers, B.A., Stark, L.W.: The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence. Presence Teleoperators Virtual Environ. **5**(3), 263–273 (1996). MIT Press
6. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. Presence **10**(3), 266–281 (2001)
7. Thomas, F., Johnston, O.: The Illusion of Life: Disney Animation. Abbeville Press, New York (1984)
8. Johansson, P., Hall, L., Sikström, S., Olsson, A.: Failure to detect mismatches between intention and outcome in a simple decision task. Science **310**(5745), 116–119 (2005)
9. Johansson, P., Hall, L., Sikström, S.: From change blindness to choice blindness. Psychologia **51**, 142–155 (2008)
10. Haake, M., Gulz, A.: Trust in Virtual Reality. (Manuscript)
11. Johansson, P., Hall, L., Gulz, A., Haake, M., Watanabe, K.: Choice Blindness and Trust in the Virtual World. Technical Report of IEICE: HIP, 107(60), 83–86 (2007)
12. Witmer, B., Singer, M.: Measuring presence in virtual environments: a presence questionnaire. Presence **7**(3), 225–240 (1998)
13. Dahlbäck, N., Jönsson, A.: Empirical studies of discourse representations for natural language interfaces. In: Proceedings of the 4th Conference on European Chapter of the Association for Computational Linguistics (1989)
14. Slater, M., Usoh, M., Chrysanthou, Y.: The influence of dynamic shadows on presence in immersive virtual environments. In: Hansmann, W., Purgathofer, W., Sillion, F. (eds.) Eurographics: Virtual Environments 1995, pp. 8–21. Springer, Vienna (1995)
15. Garau, M., Slater, M., Pertaub, D.P., Razzaque, S.: The responses of people to virtual humans in an immersive virtual environment. Presence Teleoperators Virtual Environ. **14**(1), 104–116 (2005)
16. R Core Team: R: A Language and Environment for Statistical Computing [Computer Software]. R Foundation for Statistical Computing. Vienna (2015)