Effects Caused by Surveyor Visualization Types on Honesty in Surveys and Interviews

Jamil Young
Graduation Thesis, November 2016
Media Technology MSc program, Leiden University
Thesis supervisors: Peter van der Putten, Robin de Lange
jy.navyon@gmail.com

Abstract*—This paper outlines the usage of software-controlled avatars (SCAs) within surveys to analyze its efficacy towards improving data collection experience of the users. The purpose of this research was to evaluate the level of honesty displayed by survey respondents while using different types of visualization techniques, as social desirability factors can cause distortions and errors within data collected through surveys and interviews. To determine the influence of three different surveyor types on respondents` desire to give socially desirable answers in surveys and interviews, a web-based automated survey was developed using Unity 3D platform. By passing a short version of a Marlowe-Crowne social desirability scale test (MCS), the individuals showed different results in three different groups subjected to three different types of surveying interfaces: (i) a text based survey, (ii) a survey incorporating software controlled avatar (SCA), and (iii) survey using questions with pre-recorded videos containing actual persons. The results indicated that there is some positive effect of using pre-recorded videos on the level of providing socially desirable answers, as these respondents revealed the lowest scores on the MCS scale throughout the duration of the experiment. In addition to the MCS scale questions, other questions related to demographic details, such as (age, gender, and some open questions) were also asked to compare data from different age and gender groups. The results revealed that avatar-based surveys had the highest average score value, while text-based surveys had the lowest average score value based on MCS scale. The results showed that individuals taking avatar-based surveys were much more inclined towards impression management, leading to lower level of honesty in self-administered surveys.

***Index Terms*— Social desirability, automated survey, software controlled avatars (SCAs).**

I. INTRODUCTION

In present times, with the increasing development of virtual reality technologies, the use of computer-animated, software-controlled avatars (SCAs) and virtual humans (virtual entities that look and act similar to actual human beings, but exist in the digital world) have found their application in many spheres of online human social activity, such as gaming [1] (e.g. Second Life, World of Warcraft), online shopping [2] [3], security screening interviews [4], health care settings [5], using game environments as research laboratories [6, 7], facilitation of online fraud-related learning [8-9] as well as learning and teachers' training within the educational environment [10-11], actual feedback-based avatar control [12-13], use of avatars within a rich and immersive virtual environment (IVE) [14-15]. Avatars and virtual humans (VHs) have also been deployed in order to conduct online surveys and interviews for scientific purposes, while significantly reducing time and labor constraints [16-17]. The term *avatar*, derived from the Sanskrit word *avatarah* (e.g. descent of Hindu god from the spiritual realm into the world of human beings) has been used within numerous science fiction works for as far back as mid-1970s, until it was adopted by the scientific community later on [18]. Within the present context, avatars are termed as "online manifestations of self in a virtual world" [19] that provide "the opportunity to engage in surreal and imaginary experiences that transcend the actual world" [20]. The use of SCAs within the online IVE allow social scientists with ample opportunities to observe underlying social and psychological phenomena within social interactions in a relatively controlled and naturalistic environment [21]. IVEs provide a balance between ecological validity and experimental control settings within a highly-controlled environment, which has been used in previous research to study a wide-array of different sociological phenomena that include behavioral modeling [22], behavioral mimicry [23], non-verbal behavior [24], social facilitation and inhibition [25-26] as well as stereotyping and prejudice [22].

Perceived confidentiality and anonymity remain some of the most critical aspects of the survey research [27-29], along with the emphasis on the respondents' behavior and reaction to the request for disclosure of personal and/or sensitive information [30-32]. Consequently, assessment measures based on self-reporting surveys have been repeatedly highlighted to contain *response biases*, which are termed as individual tendencies that lead to interference in accurate self-reporting [33]. There is a long history of criticism with regards to the efficacy of self-reported measures, due to the presence of a host of different ways in which response distortion can facilitate data misrepresentation [34]. Within the present context, the response biases that threaten the accuracy of data reporting and collection has been termed as *socially desirable responding* (SDR); it is a phenomena that based on the desirable elements within the socio-cultural values, norms as well as personal traits, attitudes, behaviors and opinions, force respondents to answer in a manner that projects a positive perception in front of the reviewers [32-33].

The need for deception in reporting can arise within a wide range of different circumstances and contexts. For example, previous researches have shown that within health and personal social contexts, respondents are often reluctant to respond in an

honest manner [4]. Fear of self-disclosure can prevent individuals from openly disclosing the truth and perception management can force individuals to select socially desirable options rather than highlighting the true preferences. For example, in healthcare settings, patients are particularly afraid to disclose personal information, which can be potentially sensitive or perceived as stigmatizing in nature [5]. Within the healthcare setting, patients are increasingly susceptible to omitting or reconfiguring sensitive or personal potentially stigmatizing information in order to portray a socially desirable albeit false outlook of their medical and personal history to medical professionals [5, 6]. However, omitting critical pieces of information from health professionals in order to undertake impression management can have serious and damaging consequences for the patients as well as healthcare professionals, ranging from misdiagnosis or even severe health complications in different surgical or non-surgical procedures [5]. Similar to the healthcare setting, security screening and clearance context provides another instance in which the authenticity of the collected information through survey is paramount [4]. In this respect, it has been postulated that virtual humans (VHs) within the IVEs have adequate social skills [38, 39] to foster *rapport* (the element of harmony and synchronicity within the social interaction), while providing a safe environment in which respondents can disclose sensitive personal information [5, 40].

The main goal of this paper is to evaluate the effect of different visualization techniques on respondents` answers, and to find what kind of survey visualization types reduce the social desirable responding tendencies in data collection process. For this purpose, we have developed a SCA in order to compare responses from the following three types of survey visualizations modes: (i) a text based survey, (ii) a survey incorporating software controlled avatar (SCA), and (iii) survey using questions with pre-recorded videos containing actual persons.

The paper is divided into seven main sections. Section I is based on introductory background information, which is the foundation on which the proceeding sections are heavily dependent, in terms of the importance and relevance of the present research. Section II highlights some of the existing works in the relevant research area of avatars and VHs and their relevant research contributions to their respective research areas. Section III describes the research methods that allow us to measure the social desirability in the data collection process, and the experimental conditions of this research. In section IV, the specific details of the experiment have been outlined, which include the number of participants as well as their specific age and gender distribution information. Section V analyzes the results of the surveys and analyzed them in order to gain critical insight into the social desirability scores and their relationship with the type of survey visualization technique used. Section VI discussed the prominent findings and section VII concludes the overall research and proposes some potential avenues for future research.

II. RELATED WORKS

The use of avatars and virtual humans can be traced back to as far back as Boeing's 'First Man', which was developed in 1959 for conducting ergonomic analysis inside Boeing 747 [41]. Ever since that time, avatars and VHs are increasingly being used for a wide-range of different applications within the virtual environment in order to perform a host of different functions and within numerous applications. At present, some of the numerous applications of VHs and avatars include simulation-based learning and training, virtual actors in movies, surgery, prosthetics and rehabilitation, distance learning, rehabilitation for psychological and social issues, industrial environment simulation, computer games, product and equipment designing and maintenance, to name a few areas of usage [42]. Past researches have also shown that SCAs can find its use in many fields of human social activity. In this manner, Pollina and Barretta [4] were able to reveal that national security interviews conducted using CGAs allowed respondents to effectively engage with the agents and disclose critical information in the process; thereby validating the usage of SCAs in security clearance interviews in the future. Meanwhile, previous studies have also attempted to explore the efficacy of avatars for conducting interviews in virtual environments by replacing the human interviewers [1]. However, further research is warranted in order to truly investigate the importance of using SCAs in online and offline environments in order to validate their overall effectiveness in different scenarios (e.g. casual conversation, job interview, security clearance, psychological evaluation and counseling as well as healthcare setting for gathering background knowledge on patients' history) in comparison with their human counterparts.

When comparing computer-administered surveys and interviews conducted by humans, the former have a distinct advantage in terms of gathering important information from respondents. The two critical factors that favor computer-administered surveys include the level of anonymity guaranteed to the respondents (the aspect of "mere presence" of another person within the face-to-face interview setting) and the lack of judgement that the respondents perceive due to absence of other individuals (the element of "mere belief" of observation and judgment by another human entity), who can judge the respondents in a negative or critical manner, based on the level of sensitivity and personal nature of information being shared [43]. Consequently, when participants interacted via a computer-mediated system (i.e., in different rooms), they felt a greater sense of anonymity that lead to increased frequency of information disclosure, as compared to participants who interacted with each other in face-to-face interview settings [35]. Within the research conducted by Turner et al. [44], the aspect related to perceived judgment was minimized by allowing respondents to answer questions on separate sheets of paper that were sealed in an envelope.

Previous research has revealed that the desirability of the response has been influenced by the level of trust of respondents, as well as interviewers' perceived gender, race and culture [45-47]. Murphy et al. [48] conducted interviews in *Second Life* using avatars with different physical characteristics to evaluate

the variations in findings of the respondents, who were also represented as avatars. Despite the lack of statistically significant results, researchers revealed that physical appearance of avatars affected the response of the interviewees [48]. For example, respondents being interviewed by an obese avatar were less likely to believe that they were attractive, while also reporting a high body-mass index and lower tendency to exercise in real life [48]. Another research proposed conducting online interviews using a specialized form of avatar known as embodied conversational agent (ECA), which could receive audio and visual feedback from the respondents, allowing the ECA to adapt and modify its responses based on the received feedback [13]. As a result, the researchers were able to conclude that face-to-face interview settings fostered rapport between interviewers and respondents, due to which, they were able to engage in casual conversations that could allow interviewers to adequately assess the present psychological state of the interviewees [13].

The use of customizable avatars within mobile devices was proposed by Savel et al. [49] in order to provide a better insight into the health and medical prevention practices for HIV, which remains a sensitive topic, especially for male youth engaging in sexual activity with other male individuals. However, the study failed to adequately shed light on the overall success and efficacy of the proposed model for online self-interviewing [49]. Another research evaluated the level of engagement provided by interviews using avatars with different level of capabilities, namely their level of dialog capabilities and varying degree of facial animation [50]. The results revealed that facial animation had no real effect on the response accuracy, while dialog capability had a positive effect on the overall engagement, accuracy and tendency to request clarification in ambiguous situations by the respondents [50].

III. METHOD

Recent technologically innovative trends have greatly revolutionized the manner in which data is collected. Some of the different forms of data collection modes being practically deployed include audio-based computer assisted self-interviewing (ACASI), video-based computer-assisted interviewing (VCASI), interactive voice response system (IVRS) and virtual assisted self-interviewing (VASI), to name a few [49, 51]. Computer assisted self-interviewing is a widely employed procedure for data collection [11]. There have also been attempts to implement automated interview programs into messengers, such as ICQ [12]. In order to evaluate the different types of surveys, an experiment was setup, which required respondents to complete an online survey. Participants were only informed that they were taking part in an online anonymous survey, which evaluated different survey interfaces.

A. Prototype

A web-based automated survey was developed using Unity 3D. To make it accessible in most compatible browsers, it was deployed using WebGL which renders interactive 3D computer graphics (Fig.1) This makes it easily accessible and allows participants to carry out the survey from a remote location. In order to synchronize the wording of the questionnaire with the SCA, lip-syncing techniques were employed by recording audio files and mapping them to the model to give the best realistic appearance. Using Lip Sync Pro, a Unity3d editor extension the phonemes of each audio file was mapped to the facial and mouth 3d model.

Fig.1 A software controlled Avatar

The model of SCA was able to speak the words of the questionnaires, while the overall body movements were not included in the engagement of the avatar with the user. However, in order to give a natural feel, the avatar could move during periods of inactivity in order to give a more realistic and natural feel to the respondents. Features of the character such as the clothing, facial expression and posture was kept very neutral, in this way the avatars appearance would have low influence on the answers given by respondents. Only a female character was used with the assumption that a female character would be more pleasant for the interaction. Both the text-based survey and avatar were developed using Unity 3D. Due to video playback limitations in the WebGL platform, the video survey was developed in standard web technologies using PHP and JavaScript and embedded videos. All surveys were connected to a backend API which they retrieved the questions, and saved the answers to the database. Apart from the three survey types, a dashboard was developed to analyze and visualize the data which had been collected through the experiment. The dashboard was used to pre-calculate stats such as averages and standard deviations of all three survey types. The type of survey used within this research was based on random sampling, due to which, the participants were able to conduct the survey using one of the three visualization techniques without the ability to select based on their preferences. In this way random sampling for each type of survey could be achieved, as each respondent was randomly assigned one of the three possible modes of data collection.

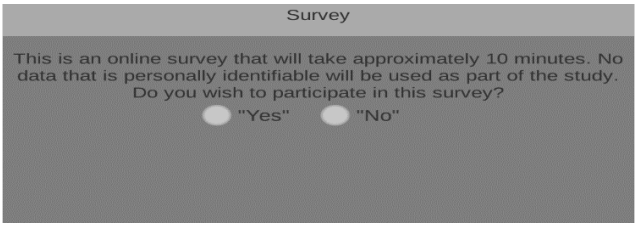

Fig.2 Text-Survey in WebGL

B. Questionnaire

During an interpersonal interaction, individuals can tend to be highly conscious regarding the manner in which the other individuals perceive and judge them, due to which individuals tend to manage personal reputation by responding in a socially desirable manner. In this respect, Marlowe-Crowne Scale (MCS) and its compact variant with fewer questions, namely the MCS-13 is typically incorporated within the interviews to examine the level of distortion being employed by individuals to enhance the social desirability of their responses [52]. Within the healthcare research domain, MCS is extensively employed to detect the level of impression management and its impact on patients' responses [53, 54]. Within the present context, the survey includes a total of 20 questions, out of which 13 questions had been incorporated from the MCS, while the rest of the questions are used for gathering personal information of the respondents. Table 1, outlines the questions based on MCS and table 2 highlights the basic personal questions that are used for gathering personal information or respondents, e.g. age, gender and personal preference for survey visualization types.

	Questions
1	Is it sometimes hard, for you to go on with your work, if you're not encouraged?
2	Do you sometimes feel resentful when you don't get your way?
3	On a few occasions, you have given up on something because you thought too little of your ability
4	Were there times when you felt like rebelling against people in authority even though you knew they were right?
5	No matter who you're talking to, you're always a good listener.
6	Have there been occasions when you have taken advantage of someone?
7	Are you always willing to admit it when you make a mistake?
8	You sometimes try to get even, rather than forgive and forget.
9	Are you always courteous, even to people who are disagreeable?
10	Have you ever been disgusted, when people expressed ideas very different from your own?
11	Have there been times when you were quite jealous of the good fortunes of others?
12	Are you sometimes irritated by people who ask favors from you?
13	Have you ever deliberately said something that hurt someone's feelings?

Table 1: Marlowe-Crowne Scale Reynolds' Form C

	Additional questions
1	What is your gender
2	What is your age group
3	What's the worst thing you've done while drunk?
4	Give your impression of this survey
5	What type of survey visualization do you prefer, text-based, an animated avatar or a video of a live person?
6	Please rate this survey on a scale of 1 to 10.

Table 2: Additional Questions

Each respondent was assigned a social desirability score based on their answers to the questions within the MCS questionnaire. The social desirability scores were calculated as follows: (i) adding a point to the overall score for 'true' response for questions 5, 7, and 9; (ii) adding a point to the overall score for 'false' response for the remaining questions. Most of the questions in MCS are general in nature, such that most individuals can relate to the scenarios being portrayed. Score was added for responses to questions in which respondents were trying to portray a positive self-image. All of the questions portray scenarios with a highlighted reaction and response of respondents such as anger, rebellion, courtesy, irritation and jealousy, etc. In these scenarios, natural reactions and responses are always different from the socially desirable responses. For example, within question 7, which states: "Are you always willing to admit it when you make a mistake?", it can be seen that under normal circumstances no individual is *always* willing to admit their mistake and those who answer 'true' to this question are bound to be lying or portraying a socially desirable perception. As, most people are only sometimes willing to admit their mistakes to others, not all the time. The participant scores could vary between 0 to 13, where the score of 0 corresponds to lowest tendency towards positive impression management and score of 13 reveals the highest inclination towards impression management. The score is intended to measure how likely a respondent is towards give answers that are more socially acceptable rather than being true in their case.

IV. Experiment

A cross sectional study was carried out using a web-based automated survey to observe the differences between the three types of survey interfaces. Participants were recruited from social media networks and university forum groups. All participants were randomly assigned to any one of the three survey types, namely: (i) text-based, (ii) a software controlled avatar (SCA)-based, or (iii) a pre-recorded video-based surveys. The survey questions were kept homogeneous throughout the different types of survey types used, so that despite changes in the form of visual stimuli, the overall responses across the different survey types could be compared and analyzed in a collective fashion to highlight the differences based on the type of survey visualization modes utilized. In this manner, the study will be able to highlight the overall preferences of individuals for each type of visualization mode and the overall tendency of respondents to engage in the different survey visualizations. Specific information obtained from participants include the individual's preference for a human or computer generated interviewer, their impression from the survey, and demographic questions related to age and gender of the respondents (Fig.3).

Age group	**No. of participants**
18-24	70
25-34	28
35-44	8
45+	4
Total	110

Table 3: Information regarding age-group of respondents

All of the 110 study participants (female = 45 and male = 65) were volunteers selected with a number of different pre-requisites that included ability to read, write, and understand English language to a certain degree, along with basic computer skills and understanding of web-browser usage. Based on the findings of the highlighted demographical details given above, male participants (65%) comprised a majority group and female participants were slightly less in comparison (45%). While, most of the respondents belonged to the age group 18-24 years (63.6%), followed by age group 25-34 years (25.5%), and the age group between 35-44 years (7.3%).

V. RESULTS

This section will shed light on the overall results and attempt to analyze the findings and their significance within the context of social desirability and the performance of each type of survey visualization. A back-end dashboard was constructed, which assisted in analyzing the data by calculating average score and time values for each survey types, as well as the standard deviation values for each of the three survey types. For each of the survey types, some of the most common types of statistical measures were used to quantitatively analyze the desirability scores for assessing the feasibility of each of the survey visualization methods employed. The table given in the proceeding section outlines some of the basic statistical information related to the three types of visualization modes for conducting surveys. It can be seen that there is not a very significant difference in the scores for avatar (M = 6.65, SD = 2.86), text (M = 5.73, SD = 2.04) and video-based (M = 5.92, SD = 3.23) visualization.

Variables	Statistical Measure	Avatar	Text	Video
Score	Mean	6.65	5.73	5.92
	Median	7	5.5	6
	Mode	7	6	6
	Standard Deviation	2.86	2.04	3.23
Time	Mean	3.81	4.74	5.61
	Standard Deviation	1.76	4.10	3.68
	No. of participants	31	42	37

Table 4: Table showing score and time values for three surveys

Time was calculated from the moment a respondent started the survey, until they submitted answers to all of the questions. The time of the avatar was the lowest among the three groups, while the respondents from the video-based survey group took the longest time to completely the survey, as it demonstrated the highest values for average and standard deviation values of time. One of the potential reasons for this can be the longer time taken to load the video, which was being streamed using links from *Youtube* website. Meanwhile, the completion time for text-based survey was intermediate between avatar and video-based survey.

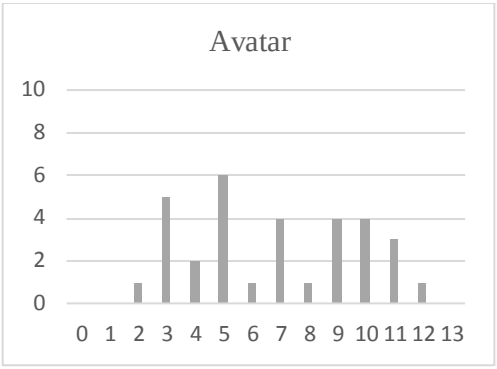

Fig. 4. Avatar survey score frequency

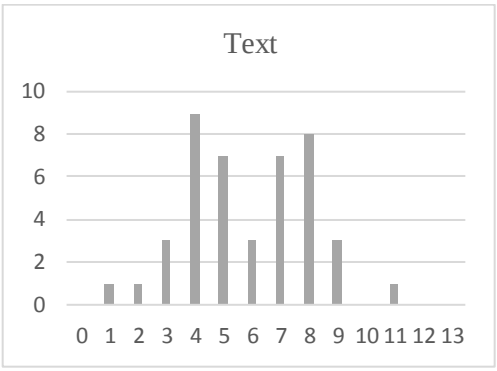

Fig. 5. Text survey score frequency

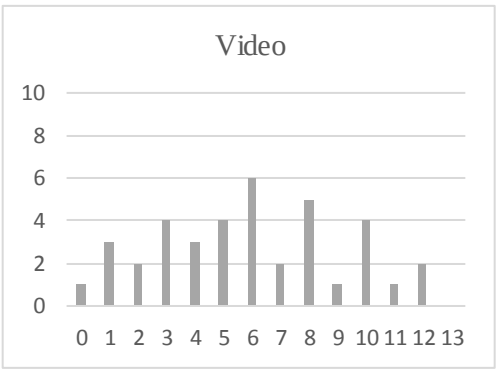

Fig. 6. Video survey score frequency

When comparing the evidences regarding the frequency and basic statistical measures obtained for the three visualization types, it can be seen that although, avatar-based surveys have the highest average score, but when comparing video and avatar with text-based surveys, the latter shows higher peak values. However, the aggregate of the average for text is less than video and avatar-based surveys raw distribution of social desirability scores of respondents who took the avatar survey has been outlined in the figure given above (Fig.4). Similarly, the distribution of social desirability scores for video-based and text-based surveys have also been highlighted in the figures

given above (Fig. 5 and 6). Standard deviation was calculated as follows:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}.$$

To gain insight into other aspects of each survey types used, the study participants were asked to choose their preferred survey visualization type. This could point to the overall preferences of the respondents; whether they are able to readily adapt to the relatively recent adoption of avatar for survey visualization or they prefer the traditionally employed text-based visualization. Figure given below sheds light on the overall level of preference of the participants in view of the three survey visualization modes highlighted in this research:

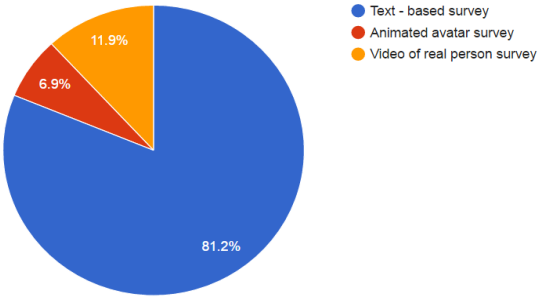

Fig. 7. Participants survey type preferences

Based on the feedback of participants, it can be seen that the overall preferences of participants ranged from highest for text-based survey to lowest for avatar visualization-based survey. In addition, participants were asked to describe their impression of the avatar-based visualization. Their answers showed that for someone it was interesting to complete the avatar survey; there were also examples of: 'not the typical survey I usually take', 'conventional (though the avatar is quite funny)' and for another respondent, 'it was scary- weird and different'. Based on the respondents' feedback, it can be inferred that the physical characteristics of the avatar could impact the preference of the users, i.e. using an avatar-based on information collected from the users before introducing them to the avatar and customizing it to fit their preferences (e.g. gender, age, and other physical characteristics such as glasses, facial features, hair, and skin tone, to name a few parameters to adjust in accordance with the personal preferences of respondents). At the same time, there are various other factors that can determine the preference for visualization type. One critical factor in this regard could be the considerable loading time or its perception for respondents, specifically in the case of the avatar application, which can force participants to favor a quicker, simpler and easily accessible text-based survey mode.

	Percentage of participants with the survey scores		
Score	Avatar	Text	Video
0	0	0	2.7
1	0	2.4	8.1
2	3.3	2.4	5.4
3	16.7	7.1	10.8
4	6.7	21.4	8.1
5	20	14.2	10.8
6	3.3	9.5	13.5
7	10	16.7	5.4
8	3.3	19.0	13.5
9	13.3	7.1	2.7
10	13.3	0	10.8
11	10	0	2.7
12	0	0	5.4
13	0	0	0

Table 5: Score distribution for the three survey visualization types

The table given above outlines the different types of visualization types and the survey scores' distribution for each of the different survey types. It can be seen from the table that each of the three visualization types have different score distributions of respondents. For example, the test scores of respondents using text-based surveys is primarily concentrated between the range 3-7, while the score range and distribution is more widely and unevenly dispersed with varying magnitude of peaks.

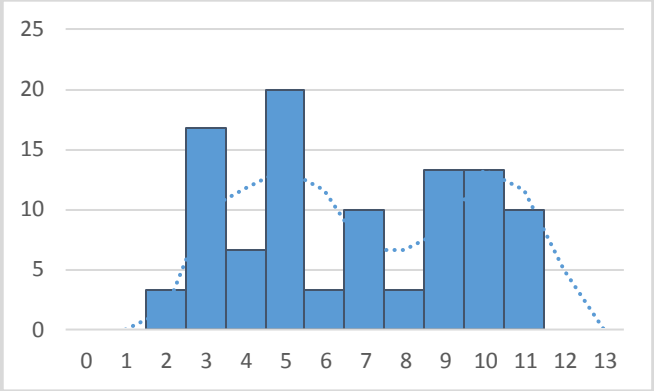

Fig. 8. Score Distribution of respondents for avatar survey

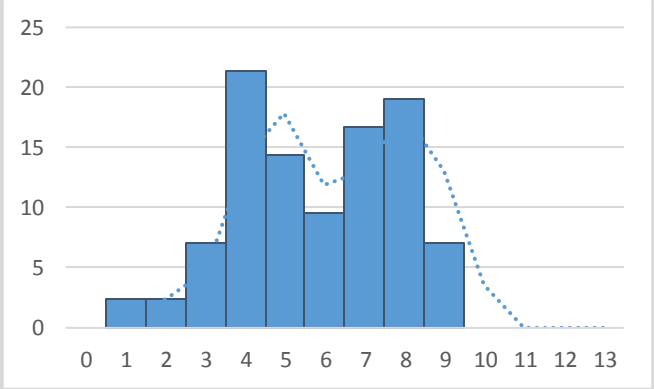

Fig. 9. Score Distribution of respondents for text-based survey

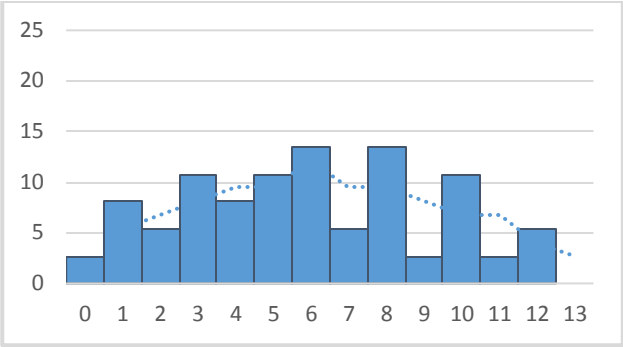

Fig. 10. Score Distribution of respondents for video-based survey

In order to properly examine the differences in score distribution between the different visualization types, there will be a need to visualize the information. Figure 8 outlines the score distribution for the avatar based survey, which shows that the highest proportion of score is concentrated between 3, 5 and 9 to 11. As a result, it can be inferred that there is a wide amount of variation in the score of the respondents of avatar-based surveys with 40% of the score concentration between 3 and 5, while 30% of score distribution between 9 and 11. The two maximum peaks in terms of the highest percentage of survey score include score 3 and 5 that contain maximum peaks of 20% and 17% respectively. However, extreme values on both ends have no share in the overall score distribution of the avatar-based survey respondents. While, the dotted line demonstrates the moving average value for each score value. Figure 9 outlines the percentage distribution of scores for text-based surveys. It can be seen that the primary percentage of score has been concentrated between 4 and 8, with the primary peaks at score 4 and 8 for more than 20% and 18% respectively. Therefore, a major percentage (~70%) of participants' score was concentrated between 4 and 8. Meanwhile, figure 10 outlines score percentages for video-based surveys. However, in comparison to the other two survey modes, video-based surveys outline a more widely distributed score percentage from the lower to higher extreme scores. Within the video-based survey, the maximum percentage of scores was 15% each for score 6 and 8. Table 9, given below highlights the pair-wise comparison of the three different visualization types in order to examine the variance between various groups. When comparing the results for each of the different survey visualization types, it can be seen that all of the three score distributions have multiple peaks at different scores. Due to the presence of multiple peaks, the distribution of scores for all three visualization types cannot be considered normally distributed in nature.

Group-wise comparison	Two-sided significance test	
	Unequal variance	Equal variance
Avatar vs. Video	0.348	0.354
Avatar vs. Text	0.165	0.141
Text vs. Video	0.803	0.797

Table 6: Results of the two-sided significance test

The table given above demonstrates the result of two-sided significance test, which were performed in order to gauge the significance of the findings. In this manner, a pair-wide comparisons is conducted between the findings of different visualization modes, so that the scores of different visualization types can be statistically analyzed in order to reveal the nature of these findings. If there findings of the pair-wise two-way t-test are statistically significant, their significance value would be less than 0.05 ($p < 0.05$). However, based on the results of all the pair-wise comparisons, it can be seen that the scores for all the survey visualization methods are not statistically significant. It can be further viewed from the results outlined in the table above, there is very little difference in the results for varying assumptions of unequal and equal variances for each of the pair-wise comparison. One of the reasons for this type of results is the small sample size allocated for each visualization type (for text-based survey n = 42, for avatar-based visualization survey n = 31 and for video-based survey n = 37) and the limited amount of variations in the score between the different groups of respondents with their different types of visualization modes.

VI. Discussion

Text surveys are the most widely used form of data collection, and due to this reason, it was revealed within this study that most of the respondents continued to prefer text based-surveys. There can be various different reasons for these preferences, ranging from slow loading speed for avatar and video-based visualization or the need of participants to save time and quickly answer all survey questions without being distracted or unnecessarily being stalled by visualizations and videos. The study revealed lowest scores (M = 5.73, SD = 2.04) for test-based surveys, which shows lowest tendency towards impression management, followed by desirability scores for video and avatar-based surveys. There were very small- level differences in the scores for each of the three. That said, analysis of the score values using two-sided significance test showed that the findings were statistically significant. Furthermore, the feedback of different respondents, along with previous researches sheds light on the importance of gender, age, and cultural disposition of avatars. In this manner, it would be interesting to develop different avatars in terms of gender and age in order to view the variations in responses and level of honesty demonstrated by different respondents. .

The findings of this research have shown that respondents of text-based survey reported lowest tendency towards impression management, while the respondents taking avatar-based surveys were most inclined towards impression management. Existing research have pointed out that avatar-based surveys were more likely to promote honesty among respondents, in comparison with face-to-face interviews [35, 43, 44], However, there is very little existing evidence suggesting any correlation or contradiction in results between text-base and avatar-based surveys. The findings of this research are interesting for various reasons: firstly, results from text-based and avatar-based surveys showed lowest and highest social desirability scores respectively and secondly, very little variations in results were revealed between text and video-based surveys. However, there is no way of proving causation in this case, as the results were statistically insignificant in nature. In order to have a better insight into the impact of different visualization types on the level of social

desirability highlighted by respondents, there will be a need for comparing the results of surveys of individual respondents and the level of variations in their response for the different visualization types, In this manner, the results of different surveys would be analyzed, not only in a collective (different respondent taking survey with same visualization types) manner, but also individually (same individual taking surveys with different visualization types) as well.

In order to further explore this phenomenon and substantiate the findings from this study, there will be a need to further explore this research in the future in order to reveal concrete and comprehensive findings that are valid and reliable in nature. It is important to understand that in the present study, the overall number of participants were unable to demonstrate a statistically significant variations from one group to another. Therefore, in the future endeavors, there will be a need to increase the number of participants for each type of visualization in order to reveal much more conclusive and statistically significant findings. For the ease of the respondents, the application could be developed for a smart phone, which could assist in increasing the overall number of respondents, due to ease of usage and accessibility. In order to explore the effects of different survey visualization types in more depth, there will be a need to experiment with other types of visualization types, such as using live surveyors instead of recorded video. Furthermore, the use of different types of avatars (i.e. male, female, young and old-aged avatars) could also shed light on the effects on perceived gender and age of avatars on honesty in surveys. Or another example could be to use an avatar that resembles someone very close to the participant (i.e. a parent) to see whether that would influence their answers. Artificial intelligence and machine learning could be incorporated within the avatars in order to give a more natural and lifelike conversation experience in order to enhance the ability to extract honest information from respondents. Furthermore, considerable difficulties were faced using the Unity WebGL platform, which is in its infancy stage. Such limitations ranged from no support for real-time text-to-speech, and lack of support for video playback in modern browsers. All the mentioned problems could be mitigated by deploying it on mobile or desktop platforms.

ACKNOWLEDGMENT

Special thanks to Peter van der Putten who took time out to supervise me, although it was very difficult to develop the prototype and conduct the research alongside a fulltime job. And for my family who always supported me consistently through every step. Also thanks to Robin de Lange for reviewing and co-supervising the paper. Lastly special thanks to all those who participated in the online survey and had enough patience to wait through the long loading times.

REFERENCES

[1] B. Hasler, P. Tuchman, D. Friedman “Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds”, Computers in Human Behavior, vol. 29, no. 4, pp. 1608–1616, 2013.

[2] M. Holzwarth, C. Janiszewski, M. M. Neumann, “The Influence of Avatars on Online Consumer Shopping Behavior” Journal of Marketing, vol. 70, pp. 19-36, 2006.

[3] N. Yee, J. N. Bailenson, N. Ducheneaut, “The Proteus effect: Implications of transformed digital self-representation on online and offline behavior,” Communication Research, vol. 33, no. 3, pp. 285-312. 2009.

[4] D. Pollina, A. Barretta, “The effectiveness of a national security screening interview conducted by a computer-generated agent”, Computers in Human Behavior, United States, vol. 39, Issue C, pp. 39-50, October 2014.

[5] G. Lucas, J. Gratch, A. King and L. Morency, “It’s only a computer: Virtual humans increase willingness to disclose”, Computers in Human Behavior, United States, vol. 37, pp. 94–100, August 2014.

[6] E. Castranova, “On the research value of large games: Natural experiments in Norrath and Camelot” Games and Culture, vol. 1, pp. 163-186, 2006.

[7] N. Ducheaneaut (2010). Massively multiplayer online games as living laboratories: Opportunities and pitfalls. In: W. S. Bainbridge (ed.). Online worlds: Convergence of the real and virtual. London: Springer.

[8] W. N. Dilla, A. J. Harrison, B. E. Mennecke, & D. J. Janvrin, “The assets are virtual but the behavior is real: An analysis of fraud in virtual worlds and its implications for the real world” Journal of Information Systems, vol. 27, no. 2, 2013.

[9] M. D. Pickard, M. B. Burns, K. C. Moffitt, “A theoretical justification for usinge modied conversational agents to augment accounting-related interviews” Journal of Information Systems, vol 27, no. 2. 2013.

[10] W. Swartout, R. Artstein, E. Forbell, S. Foutz, H. C. Lane, B. Lange, J. Morie, D. Noren, S. Rizzo, D. Traum, “Virtual Humans for Learning.” AI Magazine, pp. 13-30. 2013.

[11] C. E. Hughes, A. Nagendran, L. A. Dieker, M. C. Hynes, & G. F. Welch, (2015). Applications of Avatar Mediated Interaction to Teaching, Training, Job Skills, and Wellness. In: G. Brunnett et al. (eds.). Virtual Realities. Berlin: Springer-Verlag.

[12] J. Faller, G. Muller-Putz, D. Schmalstieg, G. Pfurtschellet, “An Application Framework for Controlling an Avatar in a Desktop=Based Virtual Environment via a Software SSVEP Brain-Computer Interface” Presence, vol. 19, no. 1, pp. 25-34, 2010.

[13] A. Mazalek, S. Chandrasekharan, M. Nitsche, T. Welsh, P. Clifton, A. Quitmeyer, F. Peer, F. Kirschner, D. Athreya. I’m in the Game: Embodied Puppet Interface improves Avatar Control. In: Proceedings of International Conference on Tangible, Embedded and Embodied Interaction (TEI), pp. 129-136. 2011.

[14] D. A. Smith, A. Kay, A. Raab, D. P. Reed. (2003). Croquet- a collaboration system architecture. In: Y. Kambayashi, S. Konomi, K. Rose (eds.). Proceedings of the First IEEE Conference on Creation Connecting and Collaborating Through Computing, pp. 2-9.

[15] F. Kuijk, K. C. A. P. Daras, B. Ravenet, H. Wei, “Autonomous Agents and Avatars in REVERIE’s Virtual Environment” Proceedings of Web3D, pp. 279-287, 2015.

[16] R. Bloomfield, “World of Bizcraft” Journal of Virtual Worlds Research, vol. 2, no. 3, pp. 3-35. 2009.

[17] R. Bloomfield, K. Rennekamp, “Experimental research in financial reporting: From the laboratory to the virtual world” Foundations and Trends in Accounting, vol. 3, no. 1, 2009.

[18] J. N. Bailenson, J. Blascovich,"Avatars" In: W. S. Bainbridge, "Encyclopedia of Human-Computer Interaction" Great Barrington, MA: Berkshire.

[19] Peterson, M., "Learning interaction in an avatar-based virtual environment: a preliminary study," PacCALL Journal, vol. 1, pp. 29-40, 2005.

[20] S. Deuchar, C. Nodder, (2003). The impact of avatars and 3D virtual world creation on learning. In: Proceedings of the 16th Annual NACCQ Conference. Palmerston North.

[21] W. S. Bainbridge, The scientific research potential of virtual worlds. Science, vol. 317, pp. 472-476. 2007.

[22] J. Fox, J. N. Bailenson, "Virtual Self-modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors" Media Psychology, vol. 12, pp. 1-25.

[23] J. N. Bailenson, N. Yee, "Digital Chameleons: Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments," Psychological Science, vol. 16, pp. 814-819, 2005.

[24] J. N. Bailenson, J. Blascovich, A. C. Beall, J. M. Loomis, "Interpersonal Distance in Immersive Virtual Environments," Personality and Social Psychology Bulletin, vol. 29, pp. 1-15. 2003.

[25] J. Blascovich, J. M. Loomis, A. C. Beall, K. Swinth, C. Hoyt, J. N. Bailenson, "Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology," Psychological Inquiry, vol. 13, pp. 103-124. 2002.

[26] C. L. Hoyt, J. Blascovich, K. Swinth, "Social Inhibition in Virtual Environments," PRESENCE: Teleoperators and Virtual Environments, vol. 12, pp. 183-195.

[27] S. Fox, & D. Schwartz, "Social desirability and controllability in computerized and paper-and-pencil personality questionnaires," Computer in Human Behavior, vo. 18, pp. 389-410. 2002.

[28] R. S. Moore, & G. M. Ames, "Survey confidentiality vs. anonymity: Young men's self-reported substance abuse," Journal of Alcohol and Drug Education, vol. 47, no. 2, pp. 32-41, 2002.

[29] T. J. Whelan & L. F. Thompson (2009). Development/validation of the PANON scale assessing survey anonymity perceptions. Paper presented at the 24th Annual Conference of the Society for Industrial and Organizational Psychology.

[30] M. P. Couper, E. Singer, R. Tourangeau, "Understanding the effects of audio-CASI on self-reports of sensitive behavior," Public Opinion Quarterly, vol. 67, pp. 385-395. 2003.

[31] A. N. Joinson, A. Woodley, U.-D. Reips, "Personalization, authentication and self-disclosure in self-administered Internet surveys," Computers in Human Behavior, vol. 23, pp. 275-285. 2007.

[32] R. Tourangeau, & T. Yan, "Sensitive Questions in Surveys," Psychological Bulletin, vol. 133, no. 5, pp. 859-883, 2007.

[33] D. L. Paulhus (2002). Socially desirable responding: The evolution of a construct. In: H. I. Braum, D. N. Jackson, & D. E. Wiley (eds.). The role of constructs in psychological and educational measurement. Mahwah, NJ: Erlbaum.

[34] A. L. Edwards (1957). The social desirability variable in personality assessment and research. New York, NY: Dryden.

[35] A. Joinson, "Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity," European Journal of Social Psychology, vol. 31, pp. 177-192, 2001.

[36] B. Farber, "Self-disclosure in Psychotherapy", Guilford, New York, NY. July 2006.

[37] M. Leary, R. Kowalski, "Impression Management: A Literature Review and Two-Component Model", Psychological Bulletin, vol. 107, No. I, pp. 34-47, 1990.

[38] J. Gratch,, N. Wang, J. Gerten, E. Fast (2007). Creating Rapport with Virtual Agents, In: Proceedings of the 7th International Conference on Intelligent Virtual Agents. Berlin: Springer.

[39] J. Gratch, S.-H. Kang, N. Wang (2013). Using social agents to explore theories of rapport and emotional resonance. In: J. Gratch & S. Marsella (eds.). Emotion in nature and artifact. Oxford: Oxford University Press.

[40] R. Hooi, H. Cho, "Avatar-driven self-disclosure: The virtual me is the actual me," Computers in Human Behavior, vol. 39, pp. 20-28, 2014.

[41] W. A. Fetter, "A Progression of Human Figures Simulated by Computer Graphics," IEEE Computer Graphics and Applications, vol. 2, no. 9, pp. 9-13, 1982.

[42] N. Magnenat-Thalmann, & Thalmann, D. (2004). Handbook of Virtual Humans. London: Wiley.

[43] J. Gratch, G. Lucas, A. KingIt's," Only a Computer: The Impact of Human-agent Interaction in Clinical Interviews", Louis-Philippe Morency Institute for Creative Technologies Waterfront Dr., Los Angeles, 2014.

[44] C. F. Turner, B. H. Forsyth, J. M. O'Reilly, P. C. Cooley, T. K. Smith, S. M. Rogers, H. G. Miller (1998). Automated self-interviewing and the survey measurement of sensitive behavior. In: M. P. Cooper, R. P. Baker, J. Bethlehem, C. Z. Clark, J. Martin, W. L. Nicholls II, J. M. O'Reilly. (eds.) Computer-assisted survey information collection. New York: Wiley.

[45] D. W. Davis, B. D. Silver, "Stereotype Threat and Race of Interviewer Effects in a Survey on Political Knowledge," American Journal of Political Science, vol. 47, no. 1, pp. 33-45.

[46] R. Tourangeau, M. P. Couper, & D. M. Steiger, "Humanizing self-administered surveys: Experiments on social presence in web and IVR surveys," Computers in Human Behavior, vol. 19, pp. 1-24.

[47] R. M. Groves, S. Presser, S. Dipko, "The role of topic interest in survey participation decisions," Public Opinion Quarterly, vol. 68, pp. 2-31. 2004.

[48] J. Murphy, E. Dean, S. Cook, M. Keating, (2010). The Effect of Interviewer Image in a Virtual-World Survey. Research Triangle Park, NC: RTI Press.

[49] C. Savel, S. Mierzwa, P. Gorbach, M. Lally, G. Zimet, K. Meyer, S. Souidi, "Web-based, mobile-device friendly, self-reported survey system incorporating avatars and gaming console techniques," Online Journal of Public Health Informatics, vol. 6, no. 2, pp.1-11. 2015.

[50] F. G. Conrad, M. F. Schober, M. Jans, R. A. Orlowski, D. Nielsen, R. Levenstein, "Comprehension and engagement in survey interviews with virtual agents," Frontiers in Psychology, vol. 6, no. 1578, pp. 1-20. 2015.

[51] M. W. Bell, E. Castranova, G. G. Wagner (2008). Virtual Assisted Self Interviewing (VASI): An Expansion of Survey Data Collection Methods in Virtual Worlds by means of VDCI, Berlin: Council for Social and Economic Data.

[52] P. Crowne, D. Marlowe," A new scale of social desirability independent of psychopathology", Journal of Consulting Psychology, vol. 24, No. 4, pp: 349-354, 1960,

[53] J. F. Edens, J. K. Buffington, T. L. Tomicic, B. D. Riley, "Effects of positive impression managemenet on the Psychopathic

Personality Inventory," Law and Human Behavior, vol. 25, pp. 235-256, 2001.

[54] P. Andrews & R. G. Meyer, "Marlowe-Crowne Social Desirability Scale and Short Form C: Forensic Norms," Journal of Clinical Psychology, Vol. 59, No. 4, pp. 483-492, 2003.