The Behavioural Impact of a Visually Represented Virtual Assistant in a Self-Service Checkout Context

Jeunese A. Payne
University of Abertay
Bell Street, Dundee, DD1 1HG
0501726@live.abertay.ac.uk

Graham I. Johnson
A T & R – NCR Labs
Dundee, Kingsway West, DD2 4SW
graham.johnson@NCR.com

Andrea Szymkowiak
University of Abertay
Bell Street, Dundee, DD1 1HG
a.szymkowiak@abertay.ac.uk

Our research investigated whether the presence of an interface agent – or virtual assistant (VA) – in a self-service checkout context has behavioural effects on the transaction process during particular tasks. While many participants claimed to have not noticed a VA within the self-service interface, behaviour was still affected, i.e. fewer people made errors with the VA present than in the voice-only and control conditions. The results are explained as reflective of an unconscious observation of non-verbal cues exhibited by the VA. The results are discussed in relation to possible behavioural outcomes of VA presence.

Psychology, Performance, Error Rates, Self-Service, Interface Agent, Virtual Assistant.

1. INTRODUCTION

Human-like interface agents have been found to positively influence user attitude and performance (Milewski, & Lewis, 1997; McBreen & Jack, 2001; Cowell & Stanney, 2005). This 'persona effect' is a concept inspired by the idea that human-likeness in an interface agent allows users to rely on natural interaction skills (e.g. interpreting expressions), making human-computer interaction easier, more efficient and emotionally satisfying (Sproull, Subramani, Kiesler, Walker & Waters, 1996; Dehn & van Mulken, 2000). It is the presence of a face in particular that facilitates interaction, signalling social identity by providing cues to human experiences such as emotion and personality (Sproull et al., 1996; Pandzic, Ostermann & Millen, 1999; Cowell & Stanney, 2005). However, an animated agent does not ensure successful interaction. One of the foremost criticisms is that human-like agents may be subject to over-attribution of human-like qualities that they do not possess, e.g., motivation (Sproull et al., 1996).

Human-like agents can also incur feelings of unease. According to Mori (1970, cited in Groom et al., 2009; MacDorman, Green, Ho, & Koch, 2009; and Ho & MacDorman, 2010) the initial positive emotional responses in perceivers to increased human-likeness of characters rapidly declines when they become 'too' human-like evoking unpleasant emotional responses (Groom et al., 2009; Ho & MacDorman, 2010). Furthermore, the offer of liberation from or help with tasks via an interface agent (herein virtual assistant, VA) is often tied to unpleasant feelings associated with loss of control and of predictability, which could lead to early rejection of a system (Cowell & Stanney, 2005). Characteristics such as attractiveness can also create positive and negative effects. Aesthetics in human-computer interaction generally increase satisfaction and perceived usability, resembling the "what is beautiful is good" stereotype in human-human interaction (Pandzic et al., 1999; Hassenzahl, 2004) – the belief that physically attractive people possess other positive attributes (Eagly, Ashmore, Makhijani, & Longo, 1991). An exception is the "what is beautiful is self-centred" stereotype – the belief that attractive people are vain or selfish (Eagly et al., 1991). Another concern is a lack of effective dynamic and reactive non-verbal behaviour in agents (Pandzic et al., 1999). Realism does not refer only to static character appearance, but animated behaviour (Groom et al., 2009). Movement is essential to social perception such as gaze analysis, with dynamic facial stimuli resulting in increased sensitivity to gaze (Nummenmaa & Calder, 2009).

Reeves and Nass (1996) claim that the social psychological phenomenon of behavioural change in the presence of others extends to human-computer interaction, especially if a computer displays human-like qualities. The purpose of this research was to assess how the presence of a female VA, in a self-service (SS) checkout context (Figure 1), affected user interaction. The aim was to determine whether the VA had a behavioural influence on engagement in a transaction.

2. FEATURES OF A VIRTUAL ASSISTANT

The impact of VA characteristics such as gender, ethnicity, non-verbal behaviour, and attractiveness on perceived likeability, trustworthiness, usefulness and ease-of-use have all been investigated (Sproull *et al*., 1996; Pandzic *et al*., 1999; Dehn & van Mulken, 2000; McBreen, Anderson, & Jack; McBreen & Jack, 2001; Cowell & Stanney, 2005; Qiu & Benbasat, 2010). Findings suggest that individuals exhibit similar social behaviour in the presence of interface agents as in the presence of real people. Fiske (1993) identifies gender as one of the top visually immediate categories by which people are judged, though findings relating to VAs are not straightforward. While people often prefer to interact with VAs of the opposite sex (Cowell & Stanney, 2005; Qiu & Benbasat, 2010), particular contexts can lead to different preferences for either gender. As in interactions between people, VA interaction may be influenced by stereotypes. If gender stereotypes influence VA acceptance, a female VA may be better received in customer service contexts than a male VA (Berry, Butler & de Rosis, 2005) – consistent with the stereotype that helping behaviour is a characteristic of femininity. Male and female VAs may also be subject to different standards of judgment. This is consistent with the shifting standards model of stereotypes, asserting that judgments about others are made relative to pre-existing perceptions of the group they belong to (Crosby, Stockdale, & Ropp, 2007).

Things like perceived attractiveness and age may moderate these judgements. For example, it can be argued that attractive females are judged in more socially desirable terms due to a single standard for beauty, largely dictated by a youthful appearance (Adams & Huston, 1975) and that attractiveness is more central to the female gender role (Eagly *et al*., 1991). Thus, a physically attractive female service worker, may be considered as competent as her less attractive male counterpart. Evidence also suggests that the female gender role portrays women as more trustworthy (Adams & Huston, 1975). According to Corritore, Kracher & Wiedenbeck (2003), trustworthiness (honesty and well-intentioned, unbiased action) is a construct, which, alongside expertise (knowledge and competence), makes up the concept 'credibility'. Trust in an interface occurs when there is ease of navigation, a professional look, transaction ease and no grammatical or typographical errors, and good use of visual design elements, as well as when it seemingly possesses predictability, dependability, usability, and reduced risk (Corritore *et al*., 2003). Interacting with a VA displaying these qualities should result in similar user experiences of trust. One question relates to whether perceived

VA trustworthiness is related to the effectiveness of the interaction. Cowell and Stanney (2005) found that a perceived non-trustworthy agent led participants to make more errors and perceive tasks as more effortful. Participants also found interaction with the agent to be more monotonous and less positive, and found the agent itself to be less likeable, intelligent, and accurate. The authors concluded that trustworthy behaviour should be supported, and in particular, that facial non-verbal cues should be credible.

Other means by which non-verbal social cues are communicated is dress code (Shao, Baker, & Wagner, 2004). McBreen *et al.* (2001), for example, found casually-dressed VAs were deemed more suitable for a cinema virtual environment whilst formally-dressed VAs were deemed more suitable for virtual banking. While all agents (smartly and casually dressed male and female VAs) were considered friendly, competent and polite, etc., their perceived trustworthiness was dependent on context. The authors suggested that responses to VAs in more serious online environments (i.e., banking and travel) could be improved by increasing perceived trustworthiness and participant confidence in the system through VA dress. This shows how non-verbal social cues (e.g. uniforms) can prompt immediate visceral affective responses even in the absence of a rational basis for trust (Riegelsberger, Sasse, & McCarthy, 2005)

The focus on non-verbal cues is part of a larger debate on the usefulness of anthropomorphism. In e-retail, users tend to enjoy conversational abilities in VAs and a majority of users prefer to see a VA *within* an interface (McBreen & Jack, 2001). Ensuring human-like characteristics (verbal and non-verbal) *complement* each other is where problems can arise (McBreen & Jack, 2001). While people may be unaware of mis-matched verbal and non-verbal behaviour, both are integrated in understanding spoken messages (Cassell, McNeill, & McCullough, 1998, cited in Isbister & Nass, 2000). Verbal and non-verbal behaviour should also be dynamic and appropriate, with steps taken to ensure it is not irritating, distracting or exaggerated (Pandzic *et al.*, 1999; McBreen & Jack, 2001).

The purpose of our study was to establish how the presence of a VA is noted in the context of a retail interaction and whether the non-verbal cues it provides improve performance, or whether information provided as text and voice, or even text-only, results in equivalent outcomes. More specifically, we assessed whether the presence of an appropriately dressed female VA in SS checkout context affected user interaction.

Figure 1: The Interface Design with VA

The aim was to determine whether a VA would have a behavioural, as opposed to subjective, effect on users engaging in a transaction.

3. PREDICTIONS

The behavioural impact of a VA in a SS context was investigated. The independent variable was 'VA Presence', with three levels: visual and verbal presence (with text); voice-only (with text); and control (text-only). There were behavioural and attitudinal dependent variables. The (behavioural) measure focussed on in this paper is error rate (incorrect button presses). It was predicted that the presence of a VA would reduce error rates such that the 'Assistant Present' condition would lead to fewer errors than the 'Voice-Only' condition which would lead to fewer errors than the 'Control'.

4. METHOD

4.1 Participants

Sixty-three participants were recruited via snowball and convenience sampling, 23 of whom were female. Twelve participants were pooled from the University of Abertay, Dundee, most of whom were students. Fifty-one participants were pooled from NCR, Dundee. One participant did not offer their age. With this exception, the age of participants ranged from 21-59 years, the mean age being 36 years. Twenty-one participants were randomly assigned to each of the three conditions.

4.2 Materials & VA Design

A participant information sheet, consent form, instruction sheet, and debriefing sheet were created. The information sheet detailed the study's purpose, who the research was supported by, what the study would entail, how long it would take, and ethical considerations. The instructions detailed the order of tasks to be performed using the SS checkout system. Participants engaged in simulated scanning, selecting, and weighing items, and entering a pin number to 'purchase' them.

A Mars™ Bar, a plastic apple labelled 'Granny Smith', three artificial croissants in a transparent plastic bag, and a packet of Polos™ were to be 'scanned' across the screen or 'selected' depending on whether or not they had a barcode. These items were chosen to provide varied tasks to complete, progressing in difficulty, i.e., simple scanning; quick selection and weighing; a more complex search task; and then dealing with an item that would fail to scan. Items were set up in the order to be 'purchased' because the transaction process was programmed such that it looked as interactive while, in fact, it was a fixed sequence of events, i.e., a Mars™ Bar would always appear on the receipt first regardless of what was 'scanned'.

The 'scanner' was a blue/grey bar – a button along the bottom of the touch screen. Scales were placed below the screen for 'weighing'. A video recorder recorded participant hands interacting with the interface. When present, the VA appeared on the left hand side. The VA was created using Autodesk Maya, which was then converted into Adobe Flash. The VA was female based on the discussed preference for female VAs. The white polo shirt was informed by a previous qualitative study investigating opinions of supermarket employee dress via semi-structured interviews (Payne, 2010, Internal Technical Report). The VA was designed to be reasonably attractive and not too distracting.

Bespoke software was developed to record interaction with the SS system, including the number of errors. For the purposes of the study, rather than allowing participants to continue down an incorrect route, incorrect button presses were logged and the screen remained until the right decision was made. If a participant indicated they were unable to determine where they were going wrong and, thus, stuck at a particular stage, the researcher moved them on to the next stage.

Figure 2: The Interface in Use

4.2 Procedure

Participants provided their age, sex, occupation, and how often, if at all, they used SS checkouts. In

the 'VA present' condition the first screen in the interface had the VA waiting next to a 'Press to Start' button. In the voice-only and control conditions, no VA was present. Following this, (i.e., pressing the button to start) participants began the transaction, consisting of four tasks. After completing these, the pin number was keyed in followed by 'Enter'.

The first task was to scan and bag a Mars™ Bar. To help participants do this, whilst saying "Please scan your first item, or look up item using the *Look up Item* key", the VA indicated where the scanner was and then the Look up Item button using non-verbal cues (i.e., looked down and across at the scanner on the screen and then the Look up Item button as she was saying it). The VA also, by default, had a slight, closed smile. Participants were asked to make sure they read the instructions carefully and were also reminded upon picking up the first item where the scanner was and what they had to do. After scanning the first item (Mars™ Bar), it appeared on the screen and participants were expected to place the item in or next to the bag provided as they preferred.

The second task was to select and weigh a 'Granny Smith' apple. Participants were prompted to scan the next item by the VA verbally: "Please scan the next item, or look up item using the *Look up Item* key" along with the non-verbal cues detailed. The popular list (of six items) emerged automatically after the 'Look up Item' key was pressed. The VA asked participants to: "Please select your item", followed by "Scales are below the scanner", also indicated non-verbally (Figure 3). Once selected, the right kind of apple was to be selected and 'weighed', after which, it appeared on the screen as part of the updated receipt and total.

Figure 3: *VA with Non-Verbal Cue*

The third item (a bag of three croissants) required a quantity to be entered on a numerical keypad. The VA prompted participants as before ("Please scan the next item, or …") with the accompanying non-verbal cues. The croissants were found by pressing 'B-C' on a list of alphabetical buttons (Figure 4) and scrolling down, following which, participant entered the number of croissants on the keypad.

Figure 4: *Items Listed in B-C category*

The final item was a packet of Polo™ mints which 'failed' to scan, responded to by the VA by raising her eyebrows to demonstrate slight surprise and embarrassment. Participants were prompted to enter the item's code verbally and non-verbally, i.e., looking at the 'Type in Code' button. The code on had to be entered correctly before continuing.

Those in the voice condition experienced the same sequence of events without the visual presence of the VA. The same items were purchased in the same order. Those in the control condition also experienced the same events, the difference being that there was no vocal or visual presence, only the instructional text of the first two conditions.

5. RESULTS

There was partial support for the hypothesis that VA presence would reduce error rates. Participants in the VA present condition made fewer errors than those in the voice-only condition who made fewer errors than those in the control. The minimum errors in the VA condition was 0 and the maximum was 9 compared to 0 and 6 for the voice-only condition, and 1 and 16 for the control condition. The data did not meet parametric assumptions, thus, a Kruskal Wallis test was performed. The VA present and voice-only conditions had the same median (1), a lower median than the control condition (3). The results show a significant difference in the number of errors made between the conditions [$X^2 = 13.649$, df = 2, p = 0.01].
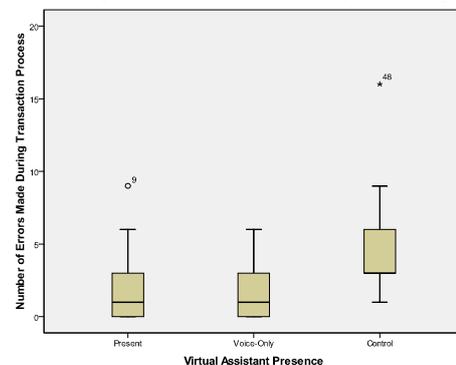

Figure 5: *Median errors and interquartile ranges for the VA present, voice-only and control conditions*

To investigate this, three follow-up Mann-Whitney U tests were carried out. On further analysis of the VA present condition and voice-only condition, a Mann-Whitney U test found the results here were not significant at the 0.017 level (U = 208.5, N1 = 21, N2 = 21, p > 0.017). On further analysis of the difference between the VA present condition and the control condition, a Mann-Whitney U test found the results were significant at the 0.017 level (U = 99, N1 = 21, N3 = 21, p = 0.002). Finally, on further analysis of the difference between the voice-only and the control conditions, a Mann-Whitney U test found the results were significant at the 0.017 level (U = 92.5, N2 = 21, N3 = 21, p = 0.001).

There are two main caveats identified with looking at the mean and median of error rates between the conditions. The first is that range of errors in the VA condition was larger (0-9) than the voice-only condition (0-6). One high error rate could have distorted how those in the VA compared to those in the voice-only condition. The second was that some incorrect responses were the same response repeated presumably because participants believed they were making the right response and that the interface was failing to register it.

Thus, the additional analysis looks at the frequency of participants who were error-free, the prediction being that VA presence will reduce error rates. This was supported. 43% of participants in the VA condition were error free compared to 29% in the voice-only condition and 0% in the control. Table 1 shows that more participants than expected were error free in the VA condition (9 v 5) than the voice condition (6 v 5). It also shows more participants than expected made errors in the control condition (21 v 16). The pattern of results suggests an association between VA presence and the frequency of participants who made errors. A Chi-Square test confirmed the results are significant. $X^2(2) = 5.025$, p < 0.01

Table 1: *Actual and Expected Errors*

Condition	Errors: Actual (Expected)		
	No. Error Free	*No. with Errors*	*Total*
VA PRESENT	*9 (5)*	*12 (16)*	*21*
VOICE-ONLY	*6 (5)*	*15 (16)*	*21*
CONTROL	*0 (5)*	*21 (16)*	*21*
TOTAL	*15*	*48*	*63(N)*

The qualitative data collected seems to indicate that participants did not appreciate, or were unaware of the VA, revealed in comments such as:

"Didn't actually really notice her … could barely describe her"
"Only really listened to the voice, didn't really look at assistant too much"

However, the *quantitative* data indicates that behaviour was still affected, i.e. fewer people made errors in the 'VA present' than the voice-only and control conditions.

6. DISCUSSION

It was predicted that a VA would reduce error rates. The initial analysis appeared to not support this, as the VA condition performed equally to the voice condition, both of which resulted in significantly fewer errors than the control condition. However, looking at the frequency of participants who were error free, it was found that none of those in the control condition were error free compared to half the participants in the VA present condition and just less than a third of participants in the voice condition. This is interesting because although many participants claimed that the VA was unnoticed or unnecessary, behaviour was still affected. This is in line with Pandzic *et al.* (1999) who found that although participants did not perceive a facial display to be particularly useful, error rates were still reduced by its presence, suggesting unconscious utilisation of the cues that facial animation provides.

Following the additional analysis it can be implied from the current investigation that the presence of a VA leads to a productive difference in error rates. Overall, the study shows a benefit to implementing a VA of this type in a SS context, even though its effects often go unrecognised by users. The results reflect unconscious observation and exploitation of non-verbal cues exhibited by a VA. Though the current study had a number of limitations (i.e., lack of interface responsiveness, lack of personal involvement in terms of choice of products to be purchased, and scanner placement), this is not a reflection of an ineffectual application of a VA in SS checkout scenarios. Rather, with consideration of some of the limitations, by providing clear cues via a VA, more effective SS checkout use could be achieved. While this study focused on one behavioural dimension, it is evident that there are other insightful measures of success. Future studies utilising other measures (objective and subjective) are planned.

7. ACKNOWLEDGEMENTS

8. REFERENCES

Adams, G. R., & Huston, T. L. (1975). Social perception of middle-aged persons varying in physical attractiveness. *Developmental Psychology,* 11, 657-658

Berry, D. C., Butler, L. T., & de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies,* 63, 304-327

Crosby, F. J., Stockdale, M. S., & Ropp, S. A. (Eds.) (2007). *Sex discrimination in the workplace.* Oxford: Blackwell Publishing

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: Concepts, evolving theories, a model. *International Journal of Human-Computer Studies,* 58, 737-758

Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human-Computer Studies,* 62, 281-306

Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies,* 52, 1-22

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin,* 110, 109-128

Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology,* 44, 155-194

Groom, V., Nass, C., Chen, T., Nielsen, A., Scraborough, J. K., & Robles, E. (2009). Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies,* 65, 842-191

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction,* 19, 319-349

Ho, C-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior,* 26, 1508-1518

Isbister, K., & Nass. C. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Interaction,* 53, 251-267

MacDorman, K. F., Green, R. D., Ho, C-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior,* 25, 695-710

McBreen, H., Anderson, J., Jack, M. (2001). Evaluating 3D embodied Conversational Agents in Contrasting VRML Retail Applications. *Proceedings of International Conference of Autonomous Agents Workshop on Multimodal Communication and Context in Embodied Agents,* Canada, 2001, 83-87

McBreen, H. M., & Jack, M. A. (2001). Evaluating humanoid synthetic agents in e-retail applications. *IEEE Transactions on Systems, Man, and Cybernetics,* 31, 394-405

Milewski, A., & Lewis, S. (1997) Delegating to Software Agents. *International Journal of Human-Computer Studies,* 46, 285-500

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences,* 13, 135-143

Pandzic, I. S., Ostermann, J., & Millen, D. (1999). User evaluation: Synthetic talking faces for interactive services. *The Visual Computer,* 15, 330-340

Payne, J. A. (2010). Evaluating self-service interface agents: A consumer perspective. *Internal Technical Report (Unpublished Document)*

Qiu, L., & Benbasat, I. (2010). Do Looks Matter? A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human-Computer studies, In Press*, 1-20

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* CSLI Publications, Cambride

Riegelsberger, J., Sasse, M. N., & McCarthy, J. D. (2005). The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies,* 62, 381-422

Shao, C. Y., Baker, J., & Wagner, J. A. (2004). The effects of appropriateness of service contact personnel dress on customer expectations of service quality and purchase intention: The moderating influences of involvement and gender. *Journal of Business Research,* 57, 1164-1176

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction,* 11, 97-124