



University of
Salford
MANCHESTER

Evaluation of human-like anthropomorphism in the context of online bidding and affordances

Murano, P and Holt, PO

| | |
|-----------------------|---|
| Title | Evaluation of human-like anthropomorphism in the context of online bidding and affordances |
| Authors | Murano, P and Holt, PO |
| Type | Article |
| URL | This version is available at: http://usir.salford.ac.uk/18985/ |
| Published Date | 2011 |

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: usir@salford.ac.uk.

Evaluation of Human-Like Anthropomorphism in the Context of Online Bidding and Affordances

Dr Pietro Murano, Prof. Patrik O'Brian Holt

Abstract— This paper presents a four condition experiment and the results concerning the wider area of investigating the effectiveness and user satisfaction of using anthropomorphic feedback at the user interface. The specific context used was online bidding. The four conditions used in the experiment were human video, human voice, human voice with anthropomorphic text and a control consisting of neutral text. The main results of the experiment showed significant differences in participants' perceptions regarding the 'humanity' of the feedback they used. As expected, the control condition consisting of neutral text incurred significantly lower ratings for the 'humanity' characteristics of the feedback. The human video condition also incurred significantly stronger perceptions regarding the appearance being human. The results were also analysed in light of the theory of affordances and the authors conclude that the four conditions used in the experiment were likely equivalent in their facilitating the affordances. Therefore the authors suggest that facilitating the affordances may be more crucial to a user interface and the users than the actual anthropomorphic characteristic of the feedback used.

Index Terms— Affordances, anthropomorphism, evaluation, user interface feedback.



1 INTRODUCTION

RESEARCHERS at various academic institutions have been investigating the use of anthropomorphism at the user interface for some time. Their work has fostered many results. However, the fundamental aspect of whether such types of feedback are better or more effective and also preferred by users is currently unresolved. Therefore the main aims of this research, which has involved several previous experiments, are to investigate the effectiveness and user preferences of anthropomorphic feedback in various contexts and to reach some concrete conclusions regarding such types of feedback in a more 'global' sense. The research is worthy and relevant because there is still no overall consensus regarding the usability of anthropomorphic interfaces.

Furthermore linked to the above points, computer scientists tend to be divided regarding the usefulness of such types of interfaces and feedbacks. Some researchers express themselves positively regarding such types of user interface basing their opinions on various studies etc, e.g. Koda and Maes [9], Maes [14], Laurel [11], Agarwal [1], Zue [31] and Takeuchi and Naito [28]. Some other researchers conversely tend to express themselves negatively concerning such user interfaces, e.g. Shneiderman and Plaisant [27].

The main structure of this paper will consist of a brief

review of some of the main literature in the anthropomorphism area then the experiment that was conducted will be described in detail. The significant results will also be presented and discussed in relation to the theory of affordances as interpreted by Hartson [7].

2 REVIEW OF MAIN LITERATURE

In this section the authors will consider some of the main literature concerning anthropomorphism being used at the user interface. While the research does span various contexts, this reflects the nature of this particular research area. However the main aim is for the reader to appreciate that there is no overall matching pattern in the results of other researchers. Furthermore, some researchers do not even attempt to determine if their anthropomorphic interface is effective or not.

In recent work by Yun and Gross [30], an anthropomorphic climate thermostat has been developed. The idea is that using a monitor-type device, a human face with different expressions is displayed to denote various settings of the thermostat, e.g. if the 'human' appears to feel cold and is wearing a scarf, this means that the current state is cold. While the concept could at first appear to be fun or intriguing, the authors do not seem to have made any attempt to measure the usability of this anthropomorphic interface. Therefore the effectiveness, user preferences and more long term attitudes towards such a product are unclear at his stage.

In a different study found in [3, 4], the effects of using flattery within feedback at the user interface were investigated. The authors wanted to find out how users would

-
- Dr Pietro Murano is with the University of Salford, School of Computing, Science and Engineering, Newton Building, Salford, M5 4WT, UK.
 - Prof Patrik O'Brian Holt is with The Robert Gordon University, School of Computing, St Andrew Street, Aberdeen, AB25 1HG, Scotland.

react to flattery by a computer. The study involved 40 [4] student participants. Each was randomly assigned to a condition. There were three conditions. The first condition involved the computer giving 'sincere praise'. The second condition involved the computer giving 'flattery (insincere praise)'. The third condition involved the computer giving 'generic feedback'. Participants in the first condition were informed that the computer feedback would be based on their performance, while participants in the second condition were informed that feedback was unrelated to their performance and that the 'evaluation' sub-system had not yet been programmed. However, in reality the feedbacks were randomly generated and the same for all three conditions. The third condition did not give participants any praise, but simply urged the participant to proceed to the next round. The actual tasks involved playing a game of 12 rounds, similar to '20-questions'. The system would request the user to think of an animal. Then the system would proceed to try and 'guess' the animal by asking questions. The participant was meant to in turn answer the questions with 'yes' or 'no'. Usually the system did not 'guess' the correct animal and therefore asked the participant to input a relevant question for future rounds. Once the participant had entered their question, the computer would provide feedback ('sincere praise' or 'flattery' etc). The study was concluded by the participants completing a questionnaire eliciting their feelings and perceptions about the interaction and computer.

The results of this experiment involved t-testing various participant scores regarding their perceptions. These were 'positive affect of user, power feeling of user, user's perception of own performance, user enjoyment of interaction, user's willingness to continue working and evaluation of computer's performance' [3]. These scores were tested for significance for each of the three conditions. There was a significant difference on the scores when the 'flattery' and 'generic feedback' conditions were compared, showing that participants in the 'flattery' condition scored their feelings and perceptions significantly higher. This was also the case when the comparison was made with the 'sincere praise' and 'generic feedback' conditions, giving similar results. No significance was found when the scores of the 'flattery' and 'sincere praise' conditions were compared. This led the authors to conclude that praise even if insincere helped users to feel more positively about themselves, their session at the computer and the computer itself.

However in a much more recent and related study by Lee [12] the results did not quite match with the study by Fogg and Nass [3]. The aim of this study was to determine if anthropomorphism aided users' placing of 'social expectations onto computers'. Linked to this, they also wanted to see if any effects would arise concerning participants' cognitive style. This involved a between-users experiment where recorded speech, artificial speech, flattering feedback, general or non-flattering feedback, low/high rationality and low/high experientiality were varied. The experiment context was a trivia game. Participants were informed that the system would generate

random responses. Therefore when participants answered the trivia questions, sometimes the system would provide (with one of the two voice types) the participant with a different answer to the one the participant had supplied. Hence the participants had to decide whether the system was correct or not and decide on a final answer to the trivia question. The main findings of this study are that the artificial speech fostered better impressions of the computer on the part of the participants. The expectation had been that the recorded (and more human) voice would have achieved better participant impressions. Also the human voice did not increase the effects of flattery towards the participants. However the participants in the human voice condition of the experiment who were designated as being less analytical (i.e. used more intuition) tended to evaluate the human voice condition more positively and they also tended to accept more the suggestions made by the system during the trivia game.

These two related studies illustrate how there is not complete agreement in results in the anthropomorphic interfaces arena. The older study by Fogg and Nass [3] had clear cut results suggesting that giving praise (even insincere praise) at the user interface was better for users. However the more recent study by Lee [12] is not so clear on this matter. Furthermore, these studies do not really indicate anything useful in terms of whether such types of anthropomorphic feedback are actually more effective and preferred by users in the context of real world tasks and scenarios.

In another study by Kramer et al [10], a prototype biometric system involving face recognition of a user was developed and evaluated. The user interface of the system was able to display the user's recognised face, an anthropomorphic face and instructions by means of voice or conventional graphical means. The system was able to take the user through the face recognition stages. This therefore involved an experiment where three methods of giving users information were varied. The three methods were a synthetic face, voice only and text only. Some of the main aspects the authors were investigating concerned 'acceptance' and 'social effects'. Overall the authors acknowledge that their results do not match completely with the work of some other researchers which suggested that the presence of anthropomorphism was more positive. In this context they quote the work of Dehn and van Mulken [2] as an example. However, the work by Kramer et al suggested that participants using the anthropomorphic entity were more 'bored' and 'less comfortable'. Also the text based condition was rated as being more 'usable' and more 'efficient' compared to all three user interface types used in the experiment.

Furthermore in a study by Yoo and Gretzel [29] in the context of a travel agency, a prototype was used which allowed participants to search for a holiday location. The system would then ask a series of questions which participants had to answer and then the system would return with a suggested location, which for experimental purposes was the same for all participants. After this interaction participants evaluated the 'credibility' and 'attrac-

tiveness' of the anthropomorphic entity. Overall the 2 x 2 experiment tested two anthropomorphic entities (human-like and an anthropomorphic suitcase - with eyes and a mouth), a human voice and no voice. The voice was used for providing encouragement to participants as they interacted with the system. The main results indicated that the human-like entity was rated as being more 'attractive'. However 'credibility' was not affected by the human-like entity. Further, the voice seemed to positively affect liking for the system, but there was no effect for 'credibility'.

These two studies also indicate that the overall results in their evaluations do not particularly match with each other. Kramer et al [10] did not find so positive results for the anthropomorphic entity in terms of boredom and comfort. However Yoo and Gretzel [29] seemed to have evidence that the anthropomorphic entity was more attractive. It could be argued that there is a link with comfort and attractiveness. However the results are not in line with each other. The reasoning here is that if one feels the anthropomorphic entity is attractive, then perhaps comfort levels should also be more positive in nature. However as stated the results do not suggest this conclusion. Also, as with the previous other studies considered above, these two studies do not show anything to indicate aspects of effectiveness and user satisfaction in useful real world contexts.

The main author of this paper has also conducted several experiments aiming to test the effectiveness and user satisfaction of anthropomorphic feedback. However similar issues arise as discussed above, where there is not a clear pattern of results to suggest a clear direction [17-22]. A further example of the author's work is found in [23]. This was a study in the context of hotel bookings. Participants needed to view some instructional material about how to carry out and cancel bookings for a hotel room and theatre performance. The between users design tested an anthropomorphic mode (animated agent) of giving the instructional content against an equivalent non-anthropomorphic mode (neutral text). The number of errors incurred, task completion success, participant hesitation and frustration and subjective opinions were measured. Overall the non-anthropomorphic condition was more effective. However the results for subjective satisfaction were inconclusive, in that the differences between the two interfaces were not statistically significant. This study also does not really match with the work by Yoo and Gretzel [29], despite being in a similar domain. Yoo and Gretzel [29] did not seem to look at effectiveness. They concentrated more on 'credibility' and 'attractiveness', which is subjective in nature. This study by the authors [23] also evaluated subjective issues which did not show significant differences in the two tested conditions. While these were not directly asking users to rate aspects of 'credibility' and 'attractiveness', they were asked many questions which were related or linked to 'credibility' and 'attractiveness'.

This therefore shows clearly that this area of research

is valuable and still requires more work to be done, because there is still overall no clear agreement in results regarding the effectiveness and user satisfaction of anthropomorphic feedback. In line with these statements, the authors present in the rest of this paper the details of an experiment conducted for the purposes of trying to further the understanding in the area of effectiveness and user satisfaction of anthropomorphism at the user interface. The experiment is in the context of online bidding. The authors used this context because it has become a common online activity with many users deriving pleasure from it. Furthermore, the results of the experiment are discussed in terms of the theory of affordances as discussed in Hartson [7].

3 ONLINE BIDDING, HUMAN VIDEO

3.1 Users

- 120 participants were recruited for this experiment.
- The participants were students of computer science. Although gender was not the main issue of this research, there were 101 male participants and 19 female participants.
- Most participants were in the 18-35 age range.
- All participants had experience or awareness of online bidding.
- The participants were all recruited from a typical computer science type university population. Specific details about the participants were then elicited by means of the pre-experiment questionnaire which principally asked a series of specific questions about bidding experience and other online activities, e.g. Do you know what an online auction web site looks like? Yes/No.

3.2 Experimental Design

A between users design was used for this experiment. Participants were randomly assigned to one of four conditions. The four conditions were human video, human voice, human voice with anthropomorphic text and a control group consisting of only neutral text. The human voice and human video consisted of a male English accented colleague.

3.3 Variables

The independent variables were (1) the types of feedback (Human Video, Human Voice, Human Voice with Anthropomorphic Text and Neutral Text) and Type of Task (Bidding on three different household items), where the values from the bids made were averaged and included in the analyses (i.e. not the tasks themselves).

The dependent variables were the participants' performance in carrying out the tasks and their subjective opinions.

The dependent measures were that the performance was measured by examining the average bid amount.

This approach was the same as the one described in Nass and Brave [24] and Huang et al [8] and is arguably suitable because it is directly related to the bidding process. From an auction point of view, the higher the bid made, the better the business outcome. From a user's perspective the lowest amount is the best outcome. Further it was decided to put the participants under pressure in the interaction by asking them to bid as quickly as possible to maximise their chance of obtaining the items. The bids and the time taken to place a bid were all recorded automatically by the prototype software. The subjective opinions were measured by means of a post-experiment questionnaire. The questionnaire had three main sections where responses were made using Likert type scales [13]. These were sections concerning the general user interface, the participants' impressions regarding the way the items were described, the participants' impressions regarding the actual items' descriptions and the participants' feelings during the interaction, e.g.

| | | | | | | | | | |
|----------------|---|---|---|---|--------------|---|---|---|--|
| Did Not Behave | | | | | Behaved Like | | | | |
| Like a Person | | | | | a Person | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |

3.4 Apparatus and Materials

- A laptop running Windows XP with 256 Mb RAM.
- The laptop's own TFT display was used - 14".
- Borland C++ Builder.

3.5 Procedure and Tasks

As indicated in the Introduction, the authors have been working in this area for some time. Therefore, the procedure and tasks used in this experiment are the same as the ones fully detailed in Murano and Holt [16]. However, the rest of the experiment is unique as it investigates different experimental conditions etc. In this section a brief summary of the tasks and experimental procedure are presented. The reader is referred to [16] for full details of the tasks and procedure.

The three tasks of the experiment were for participants to receive three different descriptions of three different household items and also view an accompanying photograph of each item. Having received the relevant items' descriptions, the participants had to then place as quickly as possible (so as to put participants under a little pressure) a single bid on each item.

The procedure involved firstly recruiting suitable participants by means of a pre-experiment questionnaire (see Users section above). When a prospective participant was deemed to be suitable for the experiment, they were contacted and an appointment was arranged for them to attend an experimental session on an individual basis. Upon their arrival, participants were asked to read a single web page which presented them with a scenario of them leaving university and obtaining their own accommodation and were therefore required to buy a few household items.

When this stage was completed, the software was launched and each item was described in one of the experimental modes (Human Video, Human Voice, Human

Voice with Anthropomorphic Text or Neutral Text), depending on the allocated condition based on random allocation for a between users experiment. When the description of the first item was completed, the participants placed and submitted a bid on the item. This was automatically recorded by the system along with the time taken to place the bid. When this was completed, the system would proceed to present the details of the second item, until all three items had been described and dealt with as described for the first item above. Having completed the three tasks, participants were then asked to complete a post-experiment questionnaire which elicited various subjective opinions (see Variables section above) about the user interface and their personal feelings during the interaction.

3.6 Results

The data were analysed using a multifactorial analysis of variance and when significance was found, the particular issues were then subjected to post-hoc testing using either a t-test or Tukey HSD test. For brevity the results for post-hoc testing are not presented in this paper, however where significance was found as shown in the tables below, the post-hoc tests confirmed significance. Lastly, for brevity, we present only the significant results in this paper.

For the variables 'Like a Person' and 'group' (i.e. Human Video, Human Voice, Human Voice with Anthropomorphic Text or Neutral Text) there is a significant ($p < 0.05$) difference. The control group (Neutral Text) significantly perceived their feedback as being less like a person than the other 3 conditions involving some human aspect of feedback. The other 3 conditions were very close in the ratings given. The F-ratio is 2.61*. This is shown in Table 1 below:

Table 1 MANOVA, Like a Person & Group (Human Video, Human Voice, Human Voice with Anthropomorphic Text and Neutral Text)

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|----------|-----|----------------|-------------|----------|
| Model | 7 | 37.64781 | 5.37826 | 2.6059 |
| Error | 112 | 231.15219 | 2.06386 | Prob > F |
| C. Total | 119 | 268.80000 | | 0.0157 |

For the variables 'Behaved Like a Person' and 'group' (i.e. Human Video, Human Voice, Human Voice with Anthropomorphic Text or Neutral Text) there is a significant ($p < 0.05$) difference. The control group (Neutral Text) significantly perceived their feedback as behaving less like a person than the other 3 conditions involving some human aspect of feedback. The other 3 conditions were very close in the ratings given. The F-ratio is 2.99*. This is shown in Table 2 below:

Table 2 MANOVA, Behaved Like a Person & Group (Human Video, Human Voice, Human Voice with Anthropomorphic Text and Neutral Text)

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|----------|-----|----------------|-------------|----------|
| Model | 7 | 54.03961 | 7.71994 | 2.9891 |
| Error | 112 | 289.26039 | 2.58268 | Prob > F |
| C. Total | 119 | 343.30000 | | 0.0065 |

For the variables 'Looked Like a Person' and 'group' (i.e. Human Video, Human Voice, Human Voice with Anthropomorphic Text or Neutral Text) there is a significant ($p < 0.01$) difference. The human video group significantly perceived their feedback as looking more like a person than the other 3 conditions. The other 3 conditions were very close in the ratings given. The F-ratio was 13.66***. This is shown in Table 3 below:

Table 3 MANOVA, Looked Like a Person & Group (Human Video, Human Voice, Human Voice with Anthropomorphic Text and Neutral Text)

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|----------|-----|----------------|-------------|----------|
| Model | 7 | 469.5080 | 67.0726 | 13.6594 |
| Error | 112 | 549.9587 | 4.9103 | Prob > F |
| C. Total | 119 | 1019.4667 | | <.0001 |

For the variables 'Sounded Like a Person' and 'group' (i.e. Human Video, Human Voice, Human Voice with Anthropomorphic Text or Neutral Text) there is a significant ($p < 0.01$) difference. The control group (Neutral Text) significantly perceived their feedback as sounding less like a person than the other 3 conditions involving some human aspect of feedback. The other 3 conditions were very close in the ratings given. The F-ratio is 9.88***. This is shown in Table 4 below:

Table 4 MANOVA, Sounded Like a Person & Group (Human Video, Human Voice, Human Voice with Anthropomorphic Text and Neutral Text)

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|----------|-----|----------------|-------------|----------|
| Model | 7 | 194.17984 | 27.7400 | 9.8816 |
| Error | 112 | 314.41183 | 2.8072 | Prob > F |
| C. Total | 119 | 508.59167 | | <.0001 |

For the variables 'Likeability' and 'group' (i.e. Human Video, Human Voice, Human Voice with Anthropomorphic Text or Neutral Text) there is a significant ($p < 0.05$) difference. The control group (Neutral Text) significantly disliked more the way the items were described, compared to the other 3 conditions involving some human aspect of feedback. The human voice was also slightly disliked, but although tending towards significance was not significant. The F-ratio is 2.22*. This is shown in Table 5 below:

Table 5 MANOVA, Likeability & Group (Human Video, Human Voice, Human Voice with Anthropomorphic Text and Neutral Text)

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|----------|-----|----------------|-------------|----------|
| Model | 7 | 27.93829 | 3.99118 | 2.2167 |
| Error | 112 | 201.65338 | 1.80048 | Prob > F |
| C. Total | 119 | 229.59167 | | 0.0380 |

3.7 Discussion of Results

As expected the participants' responses were significantly different in the questions regarding the 'humanity' of the feedback. The anthropomorphic feedbacks scored significantly higher than the other conditions, e.g. the group using the human video perceived the feedback to be significantly more like a person in appearance. Also the significant results in relation to the control group (Neutral Text) indicate the kinds of effects that one would expect, e.g. the participants in the control group perceived their feedback to be less like a person, behaving less like a person and sounding less like a person.

However the average time taken to place a bid was not significantly affected across the four conditions tested. Also the average bid was not significantly affected across the four conditions tested. Therefore effectiveness in the sense of participants significantly bidding more (good for business) or significantly bidding less (good for the customers/users) is not supported. Therefore the type of feedback did not significantly affect bidding behaviour.

Nass and his colleagues [24, 8] suggested and had results to show that bidding behaviour was affected by the type of feedback used. However our results do not bear this out. We suggest that perhaps some confounding variable was introduced in their design or procedure. While trying to be rigorous throughout the whole process, we still have retrospectively analysed our design and execution of the experiment discussed in this paper and no evident confounding variables have been identified on our part.

For user satisfaction several questions were asked by means of a post-experiment questionnaire, covering various aspects. For the general interface, questions about the ease of use, utility, satisfaction, clarity of text used, colours used and the button design being intuitive were elicited. The analysis did not show any significant differences between the four tested conditions. Participants were also asked about their impressions regarding the auction items' descriptions. The questions specifically asked opinions about the clarity of the items' description, helpfulness of the descriptions and whether the items' descriptions were understandable. For these questions, the analysis did not show any significant differences between the four tested conditions.

The questionnaire also asked several questions regarding how the participants had felt during the interaction session of the experiment. They were specifically asked about how relaxed, untroubled, comfortable, happy, ex-

cited and motivated they had felt. These questions were included to see if any of the four conditions would affect these participant 'feelings' in any particular manner. If some effect would have been observed in relation to any of the four conditions being tested, then such a result(s) would have been interesting and useful. However the analysis did not show any significant differences between the four tested conditions in relation to these questions. Therefore one can conclude that feelings and emotions are likely to be unaffected by these types of feedback in this particular context.

4 THE THEORY OF AFFORDANCES LINKED TO THE RESULTS

As discussed above, the context or domain does not satisfactorily explain the plethora of results concerning anthropomorphism. Therefore the authors of this paper are suggesting that some other explanation may be more appropriate and more closely matching the observed results. One such explanation could concern the facilitating or lack of facilitation of the affordances at the user interface whilst presenting a type of feedback to a user.

Gibson [6] was the first researcher to systematically study and propose physical affordances. As the affordances in relation to a computer user interface are different to the affordances discussed by Gibson, a detailed consideration of Gibson's theory is beyond the scope of this paper.

However, Gibson's theory of affordances has been reinterpreted for application to user interfaces. Norman [25, 26] and Hartson [7] are the main sources of the reinterpretations, with more lightweight contributions from Gaver [5] and McGrenere and Ho [15], where they started to apply affordances to computer systems and to decompose affordances into different components.

From our research it appears that Hartson [7] has made the major contribution in extending the theory to user interfaces. He identifies cognitive, physical, functional and sensory affordances. Hartson's argument is that when a user is carrying out a computing task, the users are using cognitive, physical and sensory actions. Cognitive affordances involve 'a design feature that helps, supports, facilitates, or enables thinking and/or knowing about something' [7]. Therefore in feedback design, feedback should be clear and precise. A simple example is that a button on the screen should be meaningful and related to its ultimate function. Physical affordances are 'a design feature that helps, aids, supports, facilitates, or enables physically doing something' [7]. Hartson argues that a button on the screen is a physical object acted on by a user. Furthermore the size of the button should be large enough so that a user can click it easily. This would be a physical affordance characteristic. Functional affordances concern having some purpose in relation to a physical affordance. A simple example is that usually when a user clicks on a button, there is a specific reason for that action. Lastly, sensory affordances are 'a design feature that helps, aids, supports, facilitates or enables the user in sensing (e.g. seeing, feeling, hearing) something'

[7]. Sensory affordances are linked to the earlier cognitive and physical affordances as they complement one another. This means that the users need to be able to 'sense' the cognitive and physical affordances so that these affordances can help the user.

As a reminder for the reader, the experiment discussed in this paper had four experimental conditions, i.e. Human Video, Human Voice, Human Voice with Anthropomorphic Text and a control condition consisting of Neutral Text. No significant results were observed for effectiveness and user satisfaction. The main areas where significance was observed concerned subjective opinions on the human-like characteristics (e.g. appearance etc.) of each condition and the perceptions of 'liking' a certain type of feedback, where the text only condition was significantly disliked more than the other conditions. As any good experiment will try to maintain uniformity/control across certain aspects while only varying the aspects under investigation, so the same was done with this experiment.

Therefore, with the nature of the conditions involved and the fact that the user interface was identical under each of the four conditions, with only the aspects under investigation being varied, the authors suggest that the affordances should have been the same or similar irrespective of the four different conditions. At first one could conclude that the different conditions could have affected the affordances in some way. This is because, e.g. one of the conditions was human voice only and another was text only etc. However the authors are suggesting that these differences were unlikely to be significantly affecting the results in a way as to be detected by the statistical analysis. The authors suggest that the four conditions in the experiment would have similarly facilitated the cognitive affordances as each condition amply helped the 'knowing' aspect for the cognitive affordances. Despite the fact that the four conditions did differ in the presentation mode of the information (e.g. human video and text only etc) it is suggested that in this context and given the small amount of information presented (i.e. brief descriptions of some household items) the sensory affordances would also have been virtually equally facilitated across the four conditions. This would be in terms of users 'seeing' or 'hearing' the required information to aid the placing of a bid on the items. Furthermore, although the items 'for sale' were presented using the four conditions, the actual on-screen forms used by participants to place a bid on the items, were deliberately identical across the four conditions. This would therefore mean that the physical affordances were the same across the four conditions. Linked to this, the functional affordances would also have been identical across the four conditions because the forms used for placing the bids were the same, as stated above. The form fields and buttons used were clearly labelled and were large enough to ensure unconfused and error free interaction. This is confirmed by observation during the experiment. Users did not have any problems in using the form based bidding mechanism. If some of the affordances had been violated in one or more of these four conditions, one would have expected the respective

condition(s) to have incurred in users a corresponding result, e.g. the average time to place a bid being longer etc. One result (see above) did indicate that users significantly disliked more the way the items were described in relation to the control group consisting of neutral text. It could be argued that this perhaps indicates some lack of facilitation of the affordances in this condition. However the authors would suggest that this is not the case, because this was to do with the way the items were described, i.e. the actual text. If this condition had been affected by a lack of facilitation of the affordances, one would have also expected some of the other factors to have been significantly perceived in a negative manner, e.g. the clarity of the items' description, the helpfulness of the items' description and whether the items' descriptions were understandable. None of these were significant in nature, the means for these particular factors across the four groups, were quite similar in nature. They were all rated towards the higher positive end of the Likert type scale (i.e. perceptions were not overall negative in nature). The reason for this significant result is not fully clear, however the authors suggest that perhaps users became slightly more negative towards 'something' that was purely text-based, because users today tend to expect more multimedia type interactions. They were not aware that this actually was a control condition and the text-based nature of the interaction was designed for a very specific experimental reason.

5 CONCLUSIONS AND FUTURE WORK

In this paper a new experiment has been presented in the context of online bidding. The experiment is part of a wider investigation by the authors in the area of trying to determine the effectiveness and user satisfaction of anthropomorphism at the user interface. The results of the experiment have also been analysed in terms of Hartson's rendering of the theory of affordances. The authors are suggesting that overall the main issue may be more about whether a user interface facilitates well the affordances, rather than whether an interface is anthropomorphic or not. The authors argue that the lack of significant results in this experiment in terms of effectiveness and user satisfaction are due to the four conditions approximately equally facilitating the affordances. This does also substantially agree with a previous study done by the authors [16] of this paper. Further, the affordances issue could also explain the reason for so many different kinds of results in the literature (see Review of Main Literature section above for some examples). It could be that in some of the work of others, their tested user interfaces may have in one way or another facilitated (or not) the affordances, which may have then affected the overall results. To actually gain some concrete evidence for this, one would need access to the various prototypes used by the other researchers, which is rather difficult to achieve. However the next best approach is to develop further experiments that compare anthropomorphic and non-anthropomorphic user interfaces, where the anthropomorphic condition would deliberately violate the affor-

dances and the non-anthropomorphic condition would deliberately facilitate the affordances. A further stage would invert/contrast this approach by having an anthropomorphic condition that deliberately facilitates the affordances and a non-anthropomorphic condition that deliberately violates the affordances. If the argument is correct, one would expect the results to show increased effectiveness and user satisfaction in the experimental conditions that deliberately facilitate the affordances. Significant work using this suggested approach has already started.

Overall, the work presented in this paper is useful and important for other researchers investigating user interface aspects, because it furthers the suggestion that the affordances are very important for achieving usable user interfaces and any user interface design should facilitate the affordances.

REFERENCES

- [1] A. Agarwal, "Raw Computation," *Scientific American*, 281, 44-47, 1999
- [2] D.M. Dehn, and S. van Mulken, "The Impact of Animated Interface Agents: A Review of Empirical Research," *International Journal of Human-Computer Studies*, 52, 1-22, 2000
- [3] B.J. Fogg, and C. Nass, "Silicon Sycophants: The Effects of Computers that Flatter," *International Journal of Human-Computer Studies*, 46, 551-561, 1997
- [4] B. Friedman, (Ed) *Human Values and the Design of Computer Technology*, CSLI Publications, 1997
- [5] W.W. Gaver, "Technology affordances," *Proceedings of the ACM, CHI 91, Human Factors in Computing Systems Conference*, April 27 - May 2, New Orleans, Louisiana, USA, p79-84, 1991
- [6] J.J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin Co, 1979
- [7] H.R. Hartson, "Cognitive, Physical, Sensory and Functional Affordances in Interaction Design," *Behaviour and Information Technology*, Sept-Oct, 22 (5), p.315-338, 2003
- [8] A. Huang, F. Lee, C. Nass, Y. Paik, and L. Swartz, "Can Voice User Interfaces Say 'I'? An Experiment With Recorded Speech and TTS", Unpublished Manuscript, Stanford, California, USA, 2001
- [9] T. Koda, and P. Maes, "Agents With Faces: the Effect of Personification," *Proc. of the 5th IEEE International Workshop on Robot and Human Communication*, 189-194, 1996
- [10] N.C. Kramer, G. Bente, F. Eschenburg, and H. Troitzsch, "Embodied Conversational Agents Research Prospects for Social Psychology and an Exemplary Study," *Social Psychology*, Vol. 40(1):26-36, 2009
- [11] B. Laurel, "Interface Agents: Metaphors With Character," *Software Agents*, J.M. Bradshaw, ed., MIT Press, London, p67-77, 1997
- [12] E. Lee, "The More Humanlike, the Better? How Speech Type and Users' Cognitive Style Affect Social Responses to Computers," *Computers in Human Behavior*, 26, 665-672, 2010
- [13] R. Likert, "A Technique for the Measurement of Attitudes," Columbia University Press, NY, 1932
- [14] P. Maes, "Agents That Reduce Work and Information Overload," *Communications of the ACM*, 37(7), pp. 31-40, 1994
- [15] J. McGrenere, and W. Ho, "Affordances: Clarifying and Evolving a Concept," *Proceedings of Graphics Interface*, May, Montreal, Canada, 2000

- [16] P. Murano, and P.O. Holt, "Evaluation of Anthropomorphic Feedback for an Online Auction and Affordances," *International Journal of Computer Science Issues*, Vol 8, Issue 2, March, 2011
- [17] P. Murano, "Why Anthropomorphic User Interface Feedback Can be Effective and Preferred by Users," *7th Int. Conference on Enterprise Information Systems*, - INSTICC, Miami, 2005
- [18] P. Murano, "Anthropomorphic vs Non-Anthropomorphic Software Interface Feedback for Online Factual Delivery," *Proc. of the 7th. Int. Conference on Information Visualisation*, IEEE, p. 138, London, 2003
- [19] P. Murano, "Anthropomorphic vs Non-Anthropomorphic Software Interface Feedback for Online Systems Usage," *7th ERCIM International. Workshop on User Interfaces For All*, pp. 339-349, Paris, 2002
- [20] P. Murano, "Effectiveness of Mapping Human-Oriented Information to Feedback From a Software Interface," *Proc. 24th International Conference on Information Technology Interfaces*, Cavtat, Croatia, 2002
- [21] P. Murano, "A New Software Agent 'Learning' Algorithm," *People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres*, IEE, Manchester, 2001
- [22] P. Murano, "Mapping Human-Oriented Information to Software Agents for Online Systems Usage," *People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres*, IEE, Manchester, 2001
- [23] P. Murano, A. Gee, and P.O. Holt, "Anthropomorphic Vs Non-Anthropomorphic User Interface Feedback for Online Hotel Bookings," *9th International Conference on Enterprise Information Systems*, Funchal, Madeira, Portugal, 12-16 June - INSTICC, 2007
- [24] C. Nass, and S. Brave, *Wired for Speech How Voice Activates and Advances the Human-Computer Relationship*, The MIT Press, 2005
- [25] D.A. Norman, "Affordance, Conventions and Design," *Interactions*, May-June, p.39-42, 1999
- [26] D.A. Norman, *The Design of Everyday Things*, Basic Books, 2002
- [27] B. Shneiderman, and C. Plaisant, *Designing the User Interface: Strategies for Effective Human Computer Interaction*, Pearson Education, 2005
- [28] A. Takeuchi, and T. Naito, "Situating Facial Displays: Towards Social Interaction," *Proc. CHI'95 Human Factors in Computing Systems*, pp. 450-454, 1995
- [29] K.H. Yoo, and U. Gretzel, "The Influence of Virtual Representatives on Recommender System Evaluation," *Proceedings of the 15th Americas Conference on Information Systems*, San Francisco, California, USA, Aug 6-9, 2009
- [30] R. Yun, and M.D. Gross, "The RayMatic: A Thermostat With a Face," *5th International Conference on Tangible, Embedded and Embodied Interaction*, Funchal, 23-26 January, 2011
- [31] V. Zue, "Talking With Your Computer," *Scientific American*, 281, pp. 40-41, 1999

Dr Pietro Murano is a Computer Scientist at the University of Salford, UK. Amongst other academic and professional qualifications he holds a PhD in Computer Science. His specific research areas are in Human Computer Interaction and Usability of software systems. He publishes regularly at refereed international levels and also supervises PhD level work at the University of Salford. Dr Murano is also a full member of British Computer Society, a Chartered Engineer and is FEANI registered. Further Dr Murano is a reviewer for various international conferences/journals. (www.pietromurano.org)

Prof Patrik O'Brian Holt holds a PhD as well as other academic qualifications. He is a Research Professor of Computing (Cognitive Engineering) at Robert Gordon University, Aberdeen. His research interests include interactive visualisation of complex data and cognitive modelling of human performance. Prof. Holt is the Director of the Joint Research Institute for Computational Systems which is part of the Northern Research Partnership in Engineering, a Scottish research pooling initiative.