Österreichisches Forschungsinstitut für / Austrian Research Institute for / Artificial Intelligence

TR–2010–05

Dirk Heylen, Brigitte Krenn, Sabine Payr

Companions, Virtual Butlers, Assistive Robots: Empirical and Theoretical Insights for Building Long-Term Social Relationships

● Freyung 6/6 ● A-1010 Vienna ● Austria ●

● Phone: +43-1-5336112 ●

● mailto:sec@ofai.at ●

● http://www.ofai.at/ ●

Österreichisches Forschungsinstitut für / Austrian Research Institute for / Artificial Intelligence

TR–2010–05

Dirk Heylen, Brigitte Krenn, Sabine Payr

Companions, Virtual Butlers, Assistive Robots: Empirical and Theoretical Insights for Building Long-Term Social Relationships

The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry for Science and Research and the Austrian Federal Ministry for Transport, Innovation and Technology.

Citation: Heylen, D., Krenn, B., Payr, S. (eds.). 2010. Companions, Virtual Butlers, Assistive Robots: Empirical and Theoretical Insights for Building Long-Term Social Relationships. In Trappl R. (ed.): Cybernetics and Systems 2010, Austrian Society for Cybernetic Studies, Vienna, Austria, pp. 539-570.
Also available as OFAI technical report OFAI-TR-2010-05, Austrian Research Institute for Artificial Intelligence of the Austrian Society for Cybernetic Studies, Vienna, Austria.

Table of Contents[1]

[1]Numbers in parentheses refer to the page numbers in Trappl R. (ed.). 2010. Cybernetics and Systems 2010, Austrian Society for Cybernetic Studies, Vienna, Austria.

Introduction

At the twentieth European Meeting on Cybernetics and Systems Research (Vienna/Austria, April 6 to 9, 2010), the EU project SERA (FP7-231868) organised the symposium, "Companions, Virtual Butlers, Assistive Robots: Empirical and Theoretical Insights for Building Long-Term Social Relationships", and invited submissions with the following Call for Papers:

> *Robots and agents are becoming increasingly prominent in everyday life, e.g. as companions, user interfaces to smart homes, household robots, or for lifestyle reassurance. In these roles, they have to interact with their users in a complex social world, and must build and maintain long-term relationships with them. What is considered as long-termïs not primarily a question of absolute duration. The important aspect here is that the relationship should be sustained and sustainable once the novelty effect has worn off. Are existing theories (emotion, social, psychological, behavioural) sufficient to explain what happens in this kind of interaction, and what are the required computational models to analyse and generate respective (communicative) behaviours? What types of additional (or different) theories and models would be required? What types of data would be required and what are the best methods of obtaining and analysing such data? Do we find gender differences in the ways how companions, virtual butlers and assistive robots are perceived and how long-term relations between humans and these companion technologies are built and sustained?*
>
> *We therefore call for contributions on topics including (but not limited to):*
>
> - *Theoretical and empirical research on long-term relationships of humans with humans, animals, and machines that show complex interactive behaviours*
> - *Methodology to create knowledge about interaction with companions, virtual butlers and assistive robots*
> - *(Computational) models and architectures for sustained social interaction*
> - *Case studies and good practice in user participation in the specification and development of companion, virtual butler and assistive robot technology*

The symposium was chaired on April 8, 2010 by Sabine Payr (OFAI), Dirk Heylen (Universiteit Twente), and Brigitte Krenn (OFAI). This report is an authorized reprint of the five accepted symposium papers from the proceedings of EMCSR 2010, published as: Trappl R. (ed.): Cybernetics and Systems 2010, Austrian Society for Cybernetic Studies, Vienna, 2010.

The research leading to this publication has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 231868 and from the Austrian Federal Ministry of Science and Research under its Additional Funding Programme.

Sabine Payr

Ritual or Routine: Communication in Long-Term Relationships with Companions

Sabine Payr
Austrian Research Institute for Artificial Intelligence (OFAI)*
Freyung 6/6
A-1010 Vienna
Austria
sabine.payr@ofai.at

Abstract

On the basis of data extracted from a long-term experiment with a robotic user interface, the paper discusses prerequisites and functions of interaction rituals.

1 Introduction

The challenge of building domestic companions - virtual or robotic - is to design them in such a way that they can build and maintain social relationships with their users. Experiences so far invariably point out the users' complaints that social interaction with companions, as complex as it may seem at first, becomes repetitive and boring. The researchers' conclusion has been that much more - ideally infinite - variation and variability in behaviour has to be designed for. [Bickmore et al. 2009]

On the other end of the spectrum of "household companions", we find simple devices like robotic vacuum cleaners that do not lay any claim to being social. Still, their users develop attachment to and relationships with them, without finding their single-purpose activitiy too boring or repetitive over time.

What, then, is the solution to these contradicting observations? In other words, under what circumstances do users accept what kind and degree of repetitivity? And: is repetitivity synonymous with monotony, or is there more to it? Answers to these questions could provide valuable guidelines for the design of domestic companions.

This paper first presents and discusses two sequences of video data collected during the first stage of the SERA field study (section 2). Section 3 presents and discusses concepts and theories of interaction rituals with regard to their preconditions, signs and outcomes. Section 4 discusses Interaction Ritual Theory from the point of view of emotional processes and outcomes, in particular in long-term relationships. In the concluding section 5, we will be able to formulate more concrete questons for future research on human-companion interaction.

2 Backstage and Front Performance

The two video clips that are the basis for this paper stem from the first round of data collection in the on-

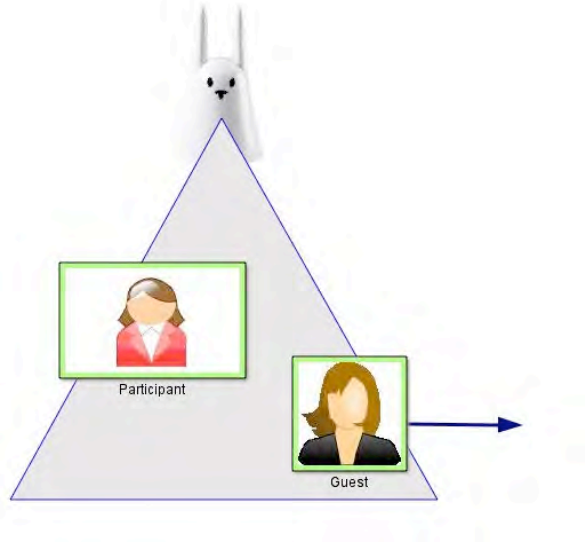

Fig. 1: Position of the participant and her guest in the second video sequence. The guest moves out of the camera's field of vision (grey triangle) during the interaction.

going EU project SERA (FP7, no. 231868). For a description of method and participants see [Klamer & Ben Allouch 2010, this volume]. The video recording was activated by the participants after a request by the

* The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry for Transport, Innovation, and Technology and by the Austrian Federal Ministry for Science and Research.

the Nabaztag (www.violet.com), a rabbit-like robotic interface. The following two sequences were recorded on two consecutive days halfway through the experiment by one of the three participants. Both were the only recordings made on that particular day. In the transcripts, N stands for Nabaztag, P2 for the participant (female, age 50+), and (in the 2nd clip) G for a (female, somewhat younger) guest of P2, probably a friend. The diagram (Fig. 1) should help the reader to understand the spatial references in the transcript. The brackets mean:
(...) movements, non-verbal expressions, notes
[...] overlaps in the dialog
{...} interaction with N through buttons, switches

Video 1: it1_p2_KOct08_1803

(P2 is alone, faces N)

1 N: Are you feeling okay after today's activities?

2 P2: (slight frown, looks at N) {presses button - doesn't work}

3 N: (pause)

4 P2: (frown, looks up, sigh)

5 N: Press the buttons to say no or yes.

6 P2: {presses button}

7 N: Are you feeling okay after today's activities?

8 P2: {presses YES button several times} (then looks up, mouth slightly open, slight frown)

9 N: (short pause) Good.

10 P2: (closes mouth, nods, slight smile)

11 N: Do you think it was the right amount of activity for one day?

12 P2: (looks at N) {presses YES button several times}

13 N: Okay, great.

14 P2: (looks half sideways, slight smile)

15 N: Keep going with the activity plan.

16 P2: (looks at N, nods, slight smile)

17 P2: (waits a little, waves her hand at N - restrained gesture)

Video 2: it1_p2_KOct09_1041

(G stands in the doorway opposite N, P2 stands beside it in profile = start position)

19 N: You are going out? Please could you press the video button

20 P2: (turns to G) (back to start pos)

21 N: on your way past? See you later.

22 P2: (turns to N, looks at N){presses button}

23 N: Recording on.

24 P2: (start pos, steps back, sottovoce) don't know why if I'm going out.

25 G: Pardon?

26 P2: (louder) don't know why if I'm going out.

27 G: Oooh?

28 P2: (steps forward, faces N, smiles) Hello, rabbit

29 P2: (bends) {puts keys on hook switch} (sottovoce) right

30 P2: (steps back out of picture)

31 N: Welcome home. If you don't mind the video recording being done, could you press the video button please.

32 G: (laughs, leaves the scene, but remains visible to P2)

33 N: Did you have a good time?

34 P2: (forward, turns to N) {presses YES button} (looks up toward camera)

35 P2: (half turns toward G) The YES button doesn't [work]

36 N: [good]

37 P2: (displays surprise): Oh ... (?rest not intelligible)

38 N: Were you doing some exercise?

39 P2: (laughs, looks back at N) {presses YES button} (looks away, smiles)

40 N: Ok, but remember that it's important

41 P2: (nods , glances at G, laughs)

42 N: to stick to the activity plan where possible

43 P2: (turns to N, stronger nod, smiles)

44 N: Don't forget to stop if you feel tired

45 P2:(grimace, strong head-shaking, smiles)

46 N: and take regular breaks.

47 P2: (strong nods)

48 P2: (grimace, steps away, toward G): what's a break

49 P2: (looks back at N, smiles): By-ye!

50 P2: (steps away)

51 G: (off: laughs)

The dialogs of the Nabaztag are scripted. The first dialog is activated after P2 has completed the final (scheduled) activity of the day. The first part of the second video shows the dialog that is activated when the participant goes out (= removes the keys from the sensitive hook) and no scheduled activity is due. P2 then puts the keys back on the hook to activate the dialog for those occasions where the participant comes home from non-scheduled activities. In content, this dialog is quite similar to that in video 1, as the Nabaz-

tag in this study is supposed to coach and monitor the subjects' physical activities.

The similarity in content allows to concentrate on the difference in P2's behaviour in the two videos. In Video 1, she responds to the Nabaztag's utterances with slight nods and smiles. We cannot be sure whether these are feedback to the utterances alone or also, partly, expressions of satisfaction that the device is functioning (after some previous technical trouble). For the first time during the study, a greeting can be observed: she waves her hand slightly in a good-bye gesture. In the second video, feedback, facial expressions and greeting are much more expressive. The greeting is now also expressed verbally, nods and head-shakes are pronounced, facial expression is exaggerated to the point of grimacing.

The main difference in the setting of the scene is the presence of a person in the second. The strongest impression one gets when watching these two videos is the contrast between the private and the public situation, or, to put it in Goffman's [1959] terms, the backstage and the front performance. Goffman noticed that the frontstage events are characterized by dramatization and idealization. Dramatization is clearly visible in this video: P2's facial expressions are more expressive than in everyday conversation, and much more than in her private interaction with the Nabaztag (Video 1). But what could be meant by "idealization"?

In Video 1, we see the private interaction between P2 and the Nabaztag. It is private in the sense that no other person is present: it is true that there is the camera recording, but it is not evident that P2 is aware of it, because her gaze and gestures are directed toward the Nabaztag and not at the camera. In this video, we see the interaction as the researchers have imagined it: a one-on-one interpersonal dialog.

In Video 2, on the contrary, the participant performs the interaction as she sees it, or more precisely: as she wants others to see it. She creates a little drama presenting what, for her, counts as a good interaction. She would have several options for this performance of and with the Nabaztag: one would be to highlight its malfunctioning, another one would be to show its stupidity. Both of these elements are present in the first part of Video 2 (line 24 and line 35), and both would involve "taking sides" with the other person and a distancing from the Nabaztag. Instead, to get more of a performance, she initiates the "coming home" sequence by putting back the key. Her position, facing halfway between the Nabaztag and her friend, indicates that, for her, there are two "others" in this interaction, and she addresses the human and the machine in turns. Her focus of attention turns more to the Nabaztag as the interaction progresses. She elaborates a dramatic "peak" in it which starts when she finds that the button this time works perfectly (line 38). She turns to the Nabaztag, and nods and shakes her haead in synchrony with the positive and negative statements of the dialog (lines 41 to 47). Compared with the first dialog, gestures are significantly longer and more expressive. Their most striking feature is that they take up and underscore the rhythm of the Nabaztag's speech. After an "aside" to her friend (line 48), she closes the interaction with a verbal greeting (line 49) - which is the only such greeting we have recorded from this participant. This idealized interaction has many elements of what has been called a "ritual" in sociological literature.

3 Interaction rituals

3.1 The ritual performance

"Two alternative conceptions of communication have been alive ... since this term entered ccommon discourse", writes Carey [2009], and goes on to explain his distinction between the transmission and ritual views of communication.

When communication is viewed as transmission, it is understood in terms of sending, receiving, and distributing information, in general in metaphors of transportation and exchange of packaged goods (cf. the "conduit metaphor" of communication, [Reddy 1979]). Whereas, in the view of communication as ritual, it is connected with terms such as sharing and participation. It reminds of the etymological relationship of the term with communion or community. "A ritual view of communication is directed not toward the extension of messages in space but toward the maintenance of society in time" (ibid. p. 15). It does not primarily serve to impart information but to express shared beliefs and emotions.

The purpose of communication is not the transmission of information but the construction and maintenance of a meaningful cultural world. Communication is a symbolic process whereby reality is produced, maintained, repaired, and transformed. Carey [2009] illustrates the difference and the necessity to reconclie both views with the "news". What the audience finds in them is not only and not even primarily information but stories on the contending forces at work in the world. "Under a ritual view, then, news is not information but drama" that invites our participation. News are not consumed for their content, but for their promise to make the reader/spectator a member in the ongoing dramas and stories.

Goffman [1967, 1981] transformed Durkheim's analysis of ritual religious gatherings [1912] into the concept of encounter which he saw as the unit of interaction, and so brought the ritual from religion into everyday face-to-face interaction. Collins [2004] has an even broader concept of ritual. Drawing on Durkheim and Goffman, he resumes the necessary ingredients of a ritual as follows:

- co-presence
- boundaries
- common focus of attention
- sharing a common mood or experience

Where Goffman saw the stereotyped sequences of talk and other gestures (used e.g. to open, close, and repair) as the defining characteristics of rituals, Collins takes his model of interaction rituals to the whole of ordinary conversation and shows that all the characteristics of a ritual can be found here. Turn-taking, for exam-

ple, can only succeed smoothly when there is an underlying rhythmic coordination. Body movements and nonverbal behaviour are synchronized in successful interaction on such a subconscious level that even brainwaves are involved. In Conversation Analysis, such phenomena have been studied under the heading of "alignment" [Bateman 2006, Branigan 2006], but the subtleties cannot be detected with its methods. Instrumental analyses of conversations have shown that synchronization is correlated with a feeling of solidarity. The participants, in this rhythmic entrainment, do not react to each other - which would be too slow - but fall into the same rhythm so that they can anticipate the "beats" of the other's talk and turn.

Such a rhythmic coordination is performed by the participant in the second video. It is "performed" in the sense that it is dramatized: nods and head-shakes are slightly exaggerated, which becomes visible in comparison with the first video. By facing the Nabaztag and thus, for a few turns, excluding her friend from the interaction, she draws the boundaries of the interaction and acts "as if" she and the device had a mutual focus of attention. In a natural conversation, the unconscious process of alignment is the work of both participants. Here, it is the human alone who does the "job" of rhythmic entrainment by adapting to the Nabaztag.

A successful interaction ritual generates shared emotions and intensifies them: beside rhythmic entrainment, there is also emotional entrainment of whatever emotions there are. The participant in the video also shows slightly exaggerated facial expressions ("grimace" in the transcript) that reinforce nods and head-shakes with agreement and rejection.

3.2 Ritual and routine

A ritual, in the everyday meaning of the word, involves stereotyped actions such as prescribed fomulas, costume, gestures, protocols. These props contribute to the core process, but they are neither necessary nor sufficient ingredients. Indeed, if a ceremony relies only on the formal rules and elements, it fails to become a ritual. Collins [2004] calls this sort of ritual "formal" and contrasts it with "natural" rituals. A formal ritual usually is repeated periodically to keep it alive. A natural ritual, on the other hand, can come off spontaneously without explicit concern, e.g. the rituals of everyday sociability such as greetings. The borders are fluid: a natural ritual can crystallize around fixed symbols whereby subsequent rituals of this kind are increasingly formalized. The difference between the two, then, is not that the natural ritual is always and completely new and spontaneous. In fact, greetings and formal politeness are strongly stereotyped and more or less formalized through repetition. Repetition can lead to routinization if the participants lose the shared focus of attention, but some repetition and take-up is necessary for rituals to confirm their symbolic value and to renew the "emotional energy" that is their outcome.

Bedtime rituals for small children are an example of interpersonal or intra-family culture. They tend to become highly repetitive in content, sequence of events, even gestures and words. Their repetitiveness and similarity come themselves to be symbols of their meaning: the order and continuity of the world into the next day is ascertained, and the monsters of the night are effectively chased and banned. With their "magic" effect they come very close to the religious rituals described by Durkheim. What distinguishes them from mere routines is their emotional outcome. With Goffman, we could say that rituals are not repeated, but re-performed.

A routine is characterized, in contrast, by the lack of focused attention. Even if it is carried out by a group, the members act on their own as individuals (e.g. on the assembly line). Rituals can decay into routines when they lose their symbolic strength, while a familiar routine can by its repetition come to symbolize continuity itself and gain the attention of the participants, and so be "celebrated" as an emotionally gratifying ritual. Routines and rituals may share repetitiveness, but are nonetheless different in the level of attention and emotional outcome. Interaction rituals could be started spontaneously, but then be carried on with more or less variation, some will decay into routines while new ones will emerge.

4 Ritual and emotion

4.1 Emotional outcomes

The ritual as a source and catalyst of emotions has a long tradition in sociology. Already Durkheim [1912/1965] described te "emotional effervescence" as the outcome of ritual gatherings. Goffman [1967] noted that feelings of solidarity emerge in the encounter. In this line, Collins [2004] says that the long-term and most important outcome of an interaction ritual is "emotional energy". Emotional energy is more enduring than the varying transient emotions that can arise in a particular situation. The gain in positive emotional energy itself is the motivation for seeking and entering into interaction rituals. A common mood or shared feeling such as joy, anger, sadness etc. are ingredients and prerequisites of the interaction ritual. The sharing and coordination of these feelings by the group reinforces this transient emotion, but this is only the short-term effect. In the long term, what remains is what he calls an "energy": the feeling of attachment to the group, of solidarity and belonging. Collins thus makes an effort to actually ground social life in everyday interaction, to show how common conversation contributes to the (re)construction of society.

Seen from the perspective of emotion research, his concept of emotional energy is so general and all-inclusive that it risks to be empty: Collins collapses the two dimensions of valence and arousal into one by putting enthusiasm, confidence and good self-feelings at one end of the spectrum and depression, lack of initiative and negative self-feelings on the other. This leads him then to link the amount of emotional energy that individuals can take away from an interaction ritual to their dominance and power [see also Collins 1990]: the more powerful they are (e.g., a group

leader), the more emotional energy they get out of the ritual. This hypothesis serves well the sociologist, in that it allows to link interaction rituals to macro-social conflicts for power [cf. Turner & Stets 2005] but does not help much in the study of everyday interaction where power differences without conflicts are the norm.

In this regard, Affect Control Theory (ACT) [Heise 2002, 2004, MacKinnon 1994] offers a more differentiated approach to the emotional outcome of interactions. It starts out recognizing different social identities (roles) that come together, with different social and affective meanings, among which their perceived power. It goes on to state that what people seek in the interaction is confirmation of their respective identities. That the successful confirmation confers a good self-feeling remains implicit, but the outcome is doubtlessly emotional. Taking, as an example, a successful conversation between a customer and a call-center agent, Collins' model cannot well explain how both sides can come away equally satisfied from such an encounter. But both customer and agent can confirm their identities which is none other than reinforcing their solidarity and bonds with their respective social groups. While Collins is concerned mainly with in-group rituals, ACT allows us thus to take the idea of interaction rituals to out-group encounters.

The first three ingredients for a ritual (see above) can be present also in such out-group interactions, but we have to ask whether persons with different identities (i.e., with different group memberships) can "share a common mood or experience". Goffman's dramaturgical approach gives us a hint to what they can have in common: the participants share the *performance* of their respective acts. They can have in common the awareness of the stage, the roles, backstage and front. With such a modified view of interaction rituals, the theory can become relevant for human-machine interaction where fundamental differences between participants are obvious.

4.1 Rituals in long-term relationships

Companions should ideally build and maintain long-term relationships with their owners. In apparent contradiction to the findings from long-term experiments with agents and robots [e.g, Bickmore & Picard 2005, see also Klamer & Ben Allouch 2010, this volume], commonsense and experience tell us that human-human relationships are far from being without repetitiveness. There are both rituals and routines, and they evolve to take up a significant part of the communication in everyday interactions. The example of the bedtime ritual is only of them. It can be safely assumed that the longer and (spatially) closer a relationship is, the more the proportion of rituals and routines in interaction will grow. People living together do not reinvent their daily interactions from scratch every morning. Cognitive economy is one factor that leads to a preference for similar situations, uncertainty avoidance another one. Where repetitive interactions can be qualified as rituals, however, they directly contribute to emotional well-being.

The role of rituals in the emotional life of long-term interpersonal relationships has not yet been studied in detail. While exchange theories [cf. Eimler et al. 2010, this volume] are based on a *trading* metaphor of emotional cost and benefit, Interaction Ritual Theory would rather be based on a *production* metaphor, because interaction rituals can generate, out of situation, co-presence and mutual attention, an emotional surplus, i.e. the feeling of belonging (bond) that is at the centre of human relationships, regardless of their content.

In the SERA field study, semi-structured interviews were conducted with all three participants [Klamer & Ben Allouch 2010, this volume]. As in comparable studies [like e.g. Bickmore & Picard 2005], subjects qualified the interaction with the companion as repetitive and rather boring. On the one hand, this result confirms that the companion is not considered a mere "machine" in the same way as, for example, a coffee-maker or vacuum cleaner: from machines, we do not expect variation. On the contrary: deviations from usual behaviour are irritating and considered as errors. On the other hand, it leaves open the question what users expect and are ready to accept in a companion. Do they indeed expect potentially endless variation, as for example from a radio or other mediating devices that do not rely on in-built content? Or would they accept a certain degree of routine which opens the possibility to develop ritual practices in which the companion is involved? Or else, would a companion that appears and behaves more machine-like, lower the expectations of variety?

We cannot answer these questions yet. We have shown, however, that more behavioural variety in companions is only one possible conclusion to draw from user feedback. There are only a handful of studies on long-term use of companions to date, so that the development of habits, routines and rituals, and the embedding of such devices into everyday practices is an uncharted area on the map. The risk is that it will remain so if researchers do not adapt their methods of inquiry and their questions to this challenge.

The SERA field study offers a unique possibility for this research because it collects observational data (video recordings of interactions over time) instead of relying on subjective data in the form of interviews or questionnaires.

5. Resume and Outlook

In the second video presented here, the participant performs a ritual with what she imagines as the ideal companion, and we can take these hints into our research agenda:

- Re-performance vs. repetition: why, how and when exactly do users notice and criticize repetitiveness?
- A certain repetitivity of behaviour is a prerequisite for the development of rituals, but not monotony. The pattern of behavioral differentiation will have to be anything from "variations over a theme" to a song with stanzas and chorus. What amount of repetitiveness is acceptable, and is it related with ap-

pearance, user expectations, and functions of the companion?

- An interaction ritual is a mutual effort and a joint action. The participant in the experiment adapts to the Nabaztag in the performance unilaterally, but would this be sufficient in a long-term relationship, or should a companion be able to contribute through (adaptive, "performing" ...) behaviour?
- Rhythmic entrainment and subverbal alignment will require speech generation which adapts, to a certain degree, to the speed, voice and beat of the individual human speaker [Suzuki et al. 2003]. It is an open question whether absolute voice qualities [Nass & Brave 2005] are more important than these (user-)relative features.
- Co-presence: is there a difference in the evolvement of interaction rituals between physically (robots) vs. virtually (agents) embodied companions?
- What social roles and how much time/space will owners give their companions in long-term everyday use? How can companions accomodate the wide variety of user attitudes? Or should owners rather be able to contribute actively to their "social configuration"?

Companions need not and should not mimic human-human relationships. They are devices that satisfy certain needs of their owners and have their uses and functions in the owner's life. When they play a role in the owner's health, well-being and independent living, however, they assume a role that goes far beyond that of, say, a vacuum cleaner, and they have to be able to maintain that role over a longer period. In this light, it becomes essential to investigate how long-term relationships are built and re-built on the micro-level of conversational interaction.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 231868 and from the Austrian Federal Ministry of Science and Research under its Additional Funding Programme.

References

[Bateman 2006] John A. Bateman: A social-semiotic view of interactive alignment and its computational instantiation: a brief position statement and proposal. In *How People Talk to Computers, Robots, and Other Artificial Communication Partners*, edited by K. Fischer: SFB/TR 8 Spatial Cognition.

[Bickmore et al. 2009] Timothy Bickmore, D. Schulmann and C. Sidner: Issues in Designing Agents for Long Term Behavior Change. CHI'09 Workshop on Engagement by Design.

[Bickmore & Picard 2005] Timothy Bickmore, Rozalind Picard: Establishing and Maintaining Long-Term Human-Computer Relationships. ACM Transaction on Computer Human Interaction (ToCHI), 59(1):21-30.

[Branigan & Pearson 2006] Holly Branigan, Jamie Pearson: Alignment in Human-Computer Interaction. In *How People Talk to Computers, Robots, and Other Artificial Communication Partners*, edited by K. Fischer: SFB/TR 8 Spatial Cognition.

[Carey 2009] James W. Carey: *Communication as Culture.* Revised edition. New York: Routledge.

[Collins 1990] Randall Collins: Stratification, Emotional Energy, and the Transient Emotions. In *Research Agendas in the Sociology of Emotions*, edited by T. D. Kemper. New York: State University of New York Press.

[Collins 2004] Randall Collins: *Interaction Ritual Chains*. Princeton, NJ: Princeton University Press.

[Eimler et al. 2010] Sabrina Eimler, Nicole Krämer, Astrid von der Pütten: Prerequisites for Human-Agent and Human-Robot Interaction: Towards an Integrated Theory. In Trappl, R. (ed.): Proceedings of EMCSR 2010. Vienna.

[Goffman 1959] Erving Goffman: *The Presentation of Self in Everyday Life*. New York: Doubleday.

[Goffman 1981] Erving Goffman: *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

[Heise 2002] David Heise: Understanding Social Interaction with Affect Control Theory. In *New Directions in Contemporary Sociological Theory*, edited by J. Berger and M. ZelditchBoulder, CO: Rowman and Littlefield.

[Heise 2004] David Heise: Enculturating Agents with Expressive Role Behavior. In *Agent Culture. Human-Agent Interaction in a Multicultural World*, edited by S. Payr and R. TrapplMahwah (NJ): Lawrence Erlbaum Associates.

[Klamer & Ben Allouch 2010] Tineke Klamer, Somaya Ben Allouch: Acceptance and Use of a Zoomorphic Robot in a Domestic Setting. in Trappl, R. (ed.): Proceedings of EMCSR 2010. Vienna.

[MacKinnon 1994] Neil J. MacKinnon: *Symbolic Interaction as Affect Control*. Albany: State University of New York Press.

[Nass & Brave 2005] Clifford Nass, Scott Brave: *Wired for Speech*. Cambridge, MA: MIT Press.

[Suzuki et al. 2003] N. Suzuki, Y. Takeuchi, K. Ishii, M. Okada: Effects of echoic mimicry using hummed sounds on human-computer interaction. Speech Communication 40: 559-573.

[Turner & Stets 2005] J. H. Turner, J. E. Stets: The Sociology of Emotions. Cambridge: Cambridge University Press.

Prerequisites for Human-Agent- and Human-Robot Interaction: Towards An Integrated Theory

Sabrina C. Eimler, Nicole C. Krämer, Astrid M. von der Pütten
University of Duisburg-Essen
Forsthausweg 2
47048 Duisburg
Germany
{ sabrina.eimler, nicole.kraemer, astrid.von-der-puetten }@uni-due.de

Abstract

Getting people to engage with robotic and virtual artifacts is easy, but keeping them engaged over time is hard and has therefore been widely ignored. Since social engagement is a complex phenomenon, the creation of long-term appealing robots/agents requires the integration of sound interdisciplinary theoretical foundations. Starting from knowledge from human-human relationships, we present an integrated theoretical framework of prerequisites for human-agent and human-robot-communication.

1 Introduction

For quite a while now, researchers have been working towards socially interactive agents and robots and have subsequently been interested in exploring the relation between humans and robots. However, both areas of research have predominantly focused on short-term interactions and effects. Recently, an increasing number of researchers have discovered long-term relationships to be important. In line with this, the SERA (**S**ocial **E**ngagement with **R**obots and **A**gents) project explores long-term relations between humans and artificial entities like robots and agents. It deals with the prerequisites for establishing and maintaining relationships between humans and agents/robots beyond an initial interaction phase. Since humans are involved in these interactions, it is necessary to know about the conditions of human-human communication in order to be able to deduce specific design guidelines for the creation of artificial characters. Against this background, we discuss a rich repertoire of different levels of interaction and configurations of relations of human communication and work towards integrating them into a coherent model. The following overview about the resulting framework introduces the concept of need to belong and, more importantly, the Theory of Mind as essential components and discusses the implications for human-machine communication.

In doing this, as a central aspect of the framework, we focus on approaches dealing with the interpersonal dimension of human encounters, reasons and antecedents for interpersonal relations as well as the rules in communication. Furthermore, the specifics of nonverbal and verbal behavior will be addressed. Sociological considerations such as ideas on the sociology of emotion [Turner and Stets, 2005] might apply in this respect too, but will not be discussed in detail here.

2 Theory Framework

The theory framework incorporates, as its core piece, the theory of need to belong [Baumeister and Leary, 1995] and the concept of Theory of Mind [ToM; Baron-Cohen, 1995; Dennett, 1987] and their appendices. We propose that the fundamental need to belong, which will be explained in more detail in the following section, serves as an anchor point for the development of long-term relationships between humans and artificial entities since it can be understood as the basic motive leading humans to establish bonds with artificial entities. Driven by this need, humans are oriented towards others, striving to relate themselves, their thoughts and feelings to their environment. In the course of this, they are likely to form a theory about their counterpart to be able to engage successfully in meaningful communication as a basis for relationships. The need to belong can thus be considered to be an essential prerequisite and starting point for assuming that humans will establish bonds with a robot/agent.

The mechanisms comprising the fundamental parts are mediated by communicative events consisting of verbal and nonverbal information. The model (see figure 1) distinguishes mechanisms bound to perception on the one hand, and on the other hand aspects of production in communicative events. Production and reception of information in a given situation can be described with the help of assumptions derived from general systems theory. It can be assumed that perceived verbal and nonverbal information underlies the rules of Watzlawick´s five axioms [Watzlawick et al., 1969] in the way that, for example, every message includes content and relationship information and therefore makes a contribution to the establishment of the relationship. More importantly, the

meaning of incoming messages is constructed against the background of personal experience as well as common ground information – taking into account the other´s perspective.

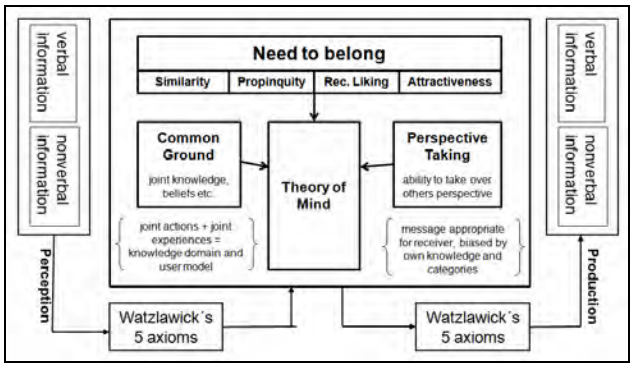

Figure 1: Theory Framework

Similarly, the sender of the message considers the other´s perspective to form a message. Thus, the outgoing message is built on the sender´s ToM about the receiver. This process will be further explained in the following sections.

It is important to consider that the following illustrations of the framework's components can be distinguished into two categories that are guided by two central questions: Which findings from human-human communication can be made use of for the design of robots? Which characteristics and capabilities are essential to a long-term engaging robot? As a first approach one can think of giving the robot certain features and characteristics that will, according to the theories and findings discussed, lead to attraction. This may for example be physical attractiveness which can be implemented before the interaction starts and is a rather static feature of the robot. Another strategy that can be followed is the idea of implementing certain theoretical assumptions enabling the robot to act autonomously, which would be the case when giving the robot ToM capabilities.

3 Factors Assisting to Establish Relationships in Human-human Communication

Being fundamental to human nature, the need to belong can be made use of in human-robot communication as a basis for the establishment of long-term relationships. In their article on belongingness, Baumeister and Leary [1995] suggest that "human beings are fundamentally and pervasively motivated by a need to belong, that is, by a strong desire to form and maintain enduring interpersonal attachments [p. 522]." This human motivation has "multiple links to cognitive processes, emotional patterns, behavioral responses, and health and well-being" [p. 522]. Consequently, all of us are interested in having warm and positive relationships and making and maintaining friendships as key conditions for happiness [Berscheid, 1985; Berscheid and Reis, 1998].

As a deep-rooted pillar of human existence, the need to belong is thus the basis for the social orientation of human beings. In order to satisfy this need we seek company of others: we build groups (e.g. families, cliques), are interested in the other's lives and help each other just because the satisfaction of the need to affiliate makes us happy. Especially in addition to other social company or to satisfy the need to belong in older or home bound people, a robot as a companion may show to be a valuable alternative.

However, since we do not build close relationships to everyone we encounter, there obviously are a number of pre-conditions influencing with whom we affiliate and under what kind of antecedents friendship and interpersonal attraction arise. These conditions have to be considered in designing robots/agents that are likely to be engaging over a long period. These aspects are subject of the following subchapters.

3.1 Propinquity

As one of the main factors leading to interpersonal attraction researchers described the propinquity effect, the tendency to form friendships or close relations to the people we often see and interact with. The propinquity effect is often explained by the mere exposure effect [Zajonc et al., 1989], i.e. the finding that the more we are exposed to a stimulus, the more likeable it becomes. However, if we initially dislike the stimulus our repulsion grows every time we are exposed to it again [Swap, 1977].

3.2 Similarity

Propinquity does affect whom we choose to be friends with. However, we do not become friends with everyone we often encounter in our direct physical surrounding. Thus, besides proximity, research has shown that it is similarity that draws people together [Berscheid and Reis, 1998; McPherson et al., 2001]. The more a person shares our attitude and opinions the more we like him or her. The same is true for similar experiences, interests and personality aspects as factors fostering the establishment of a common ground. Similarities regarding the interpersonal communication style determine to whom we are attracted and influence our evaluation of the relationship as well as the degree of satisfaction [e.g. Duck and Pittman, 1994].

3.3 Reciprocal Liking

Since we all like to be liked, we are attracted to others who behave as if they like us. No matter if the signals are nonverbal or verbal, whether we like a person or not depends on our judgment about the extent to which the other person likes us [Berscheid and Walster, 1978; Kubitschek and Hallinan, 1998]. Liking can even compensate the absence of similarity [Gold et al., 1984]. As Curtis and Miller [1986] demonstrated, reciprocal liking might as well be the result of a self-fulfilling prophecy. People, who believed they were liked by their counterpart, generally behaved more likeable and were at the same time liked more than the

participants who believed they were disliked. However, a person's self-esteem does affect this mechanism: In contrast to people with a positive or moderate self concept, people with a negative self concept tend not to respond to the friendly behavior of others and will accordingly provoke negative reactions affirming their negative self concept [Swann et al., 1992].

3.4 Physical Attractiveness

Physical attractiveness is another important determinant for liking. In general, we tend to like physically attractive people and ascribe positive qualities to them that are not connected to their outward appearance – the "what is beautiful is good" stereotype [Dion et al., 1972]. Physical attractiveness decides on liking and excels factors like intelligence, independence or sensitivity [Walster et al., 1966]. Since what we consider to be attractive is influenced by the media, it is not surprising that we share a certain number of criteria defining what is attractive.

3.5 Social Exchange Theory

A theory that brings together the different determinants of attraction is the social exchange theory [Homans, 1961; Thibaut and Kelley, 1959]. Assuming that relationships are comparable to a marketplace where costs and benefits are exchanged according to economic principles, this theory suggests that the feeling that we have about a relation does not only depend on the evaluation of the rewards and costs, but is determined by the comparison level [Kelley and Thibaut, 1978; Thibaut and Kelley, 1959] which takes into account the expected outcome of rewards and punishments the person is likely to receive in a relationship. Furthermore, the level of satisfaction also depends on your evaluation of the comparison level for alternatives, i.e. the assumption on what one would receive in an alternative relationship.

Resulting from criticism on the social exchange theory the so called equity theory was proposed. It assumes that people are concerned about equitable relationships in which the contribution of rewards and costs made by the partners are roughly equal [Homans, 1961; Hatfield et al., 1978]. Compared to inequitable relationships, in which the partners feel uneasy about the perceived imbalance, equitable relationships are the happiest and most stable relations.

3.6 Social Exchange in Long-Term Relationships

With regard to social exchange in close relations the investment model has been developed. It suggests that in long-term relationships not only the level of satisfaction with a relationship regarding rewards and costs, comparison level and the comparison level for alternatives play a role but also the perception of what has been invested that would be lost by ending the relationship [Rusbult, 1983]. Thus, in order to be able to predict the duration of an intimate relationship one has to know about these determining factors.

3.7 Implications

As these aspects mainly deal with human-human relations, the question arises what implications can be deduced for the relation between humans and robots and agents, respectively. In the beginning it was already mentioned that there are two different approaches that can be followed for the design of long-term relationships with robots. In line with the first "track" of implementing specific characteristics and features beforehand the following implications can be deduced. As a consequence of findings about propinquity, it can be suggested that in order to make use of the mechanism of mere-exposure, the agent/robot has to be within the user's reach and clearly visible as often as possible. The robot/agent should be designed in a way that it fosters interaction with its owner. Therefore, it should ideally have some similarities with the user, when believable, it could for example look human-like and conform to ideas of attractiveness, dress and speak in similar ways or show similar habits and interests. Studies with virtual agents have already shown that agents with similar personality traits, like e.g. introverted versus extroverted [Isbister and Nass, 2000] and similar appearance [van Vugt et al., 2006] like the user were evaluated more positive and likable. In line with findings about reciprocal liking the robot should give its user the impression that it likes him or her and appreciates his or her presence since this increases the likeability of the system, as long as this is authentically implemented. Depending on the setting this may well be realized with the help of ingratiation, i.e. by praising the user. For the design of robots/agents, attractiveness should be taken into account in order to increase the likelihood for the agent to be liked and to foster human motivation to establish a relationship with the robot. Finally, the robot/agent has to be of use for the user, so that he/she might at least initially feel a balance in the relation. A user's feeling of a balance between contributions and rewards from the interaction with a robot is important for the maintenance of the relationship in the beginning. However, it is important to create equitable, balanced relationships in order not to cause a bad feeling in the users and to make the relation as stable as possible.

After this initial phase in which a give-and-take rule is applied, the user hopefully perceives his/her relation towards the agent/robot as a communal relationship, so that equal contributions become less important. Ideally, the user feels a strong bond with his robot, so that he does not consider or rejects alternatives and feels bad about ending the relationship. Besides these features that generally can be implemented once before the interaction starts, a specific model of the user and the common "history" of user and robot will be needed in order to render ongoing communication, relationship management and development successful and satisfying.

4 Theories for Mutual Understanding

To be sociable, robots and agents need a representation of users, their social and cultural background, and of interaction situations and contexts. This representation (in the broadest sense) has to integrate ToM and emotionality, situational awareness and general behavioral patterns and has therefore to be more dynamic than the previously mentioned implementation guidelines. A key element is the capacity of being aware of and being able to manage socio-emotional relationships. What is meant here can aptly be illustrated by Wittgenstein´s statement "If a lion could talk we would not understand it.", referring to the fact that it is useless to implement the ability of natural speech in robots while they are unable to understand concepts which are naturally shared by humans and are taken for granted in communicative interactions.

4.1 Common Ground

Clark [1992] describes common ground as the joint basis for communication: "Two people´s common ground is, in effect, the sum of their mutual, common, or joint knowledge, beliefs, and suppositions" [p. 93]. Common ground is the basic requirement to interact with others. When there is no common ground then no communication or understanding takes place. One has to differentiate between *communal common ground,* which is derived from obvious similarities like human nature (people use language, live in groups, and have names), and *personal common ground,* which is built during interactions by joint perceptual experiences and actions. People try to ground what they do together to avoid discrepancies. According to Clark [1992] there exist several grounding principles that vitalize common ground and help to establish mutual knowledge, e.g. the *linguistic co-presence heuristic:* people assume that anything that has been said during the course of the conversation is known to the interaction partners [Clark and Carlson, 1981]; or the *principle of closure*: people try to collect evidence that they have succeeded in performing an action. Therefore, participants of a joint action give each other subtle feedback to form the mutual belief of a successful joint action. This need for feedback has also been described with regard to human-computer-interaction: Both, "telephone buttons that do not beep when pressed or a display that does not change when an action has been taken, are confusing" [Norman, 1988, p. 56].

4.2 Perspective Taking

Social perspective taking, i.e. understanding the feelings, thoughts and motivations of others, is an essential social skill that has been stressed by many researchers. According to Krauss and Fussell [1991] the role of knowing what others know is fundamental. The lack of taking the other´s perspective can be the basis for misunderstandings and dispute. Thus, tailoring the message to the knowledge of the recipient is a prerequisite for successful communication [Krauss and Fussel, 1991]. Research has shown that by taking their addressee´s knowledge and perspectives into account when formulating messages speakers´ accuracy of assessments of others´ knowledge is fairly high but they seem to be biased in the direction of their own knowledge [see also Nickerson, 1999, see below]. Krauss and Fussell [1991] summarize that those people´s assumptions of others´ knowledge shall be deemed to be hypotheses that need to be evaluated and modified over time. During interactions conversational resources might serve as feedback to check one's own assumptions on the knowledge of others.

4.3 Imputing One´s Own Knowledge to Others

Nickerson [1999] claimed that "imputing one´s knowledge to a specific other is a *default* measure; … If one has no direct knowledge of what another, whom one is addressing, does or does not know, and little or no knowledge that would provide the basis for making inferences in this regard, the only thing left to do is to use one´s own knowledge as a default assumption as to what the other know" [Nickerson, 1999, p. 745]. The ability to impute one´s own knowledge to others is thus crucial for meaningful human-human communication. He states that "[a]n obvious starting point for building a model of what another knows is what oneself knows, or think one knows" [p. 737]. This generally useful mechanism of knowledge imputation potentially includes the risk that people´s erroneous assumption about others having the same knowledge causes communication difficulties.

4.4 Theories for Understanding Others

Theory of Mind (ToM) is the ability to see other entities as intentional agents whose behavior is influenced by states, beliefs, desires etc. and the knowledge that other humans wish, feel, know or believe something [Premack and Premack, 1995; Premack and Woodruff, 1978; Whiten, 1991]. ToM is also assumed to be fundamental to human nature: "We are 'mindreaders' by nature, building interpretations of the mental events of others and feeling our constructions as sharply as the physical objects we touch. Humans evolved this ability because, as members of an intensely social, cooperative, and competitive species, our ancestors´ lives depended on how well they could infer what was on one another´s minds" [Toby and Cosmides, 1995, p. XIII]. Baron-Cohen [1995] sums up that mindreading is useful because "aside from decoding the referent of each word (computing its semantics and syntax), the key thing we do as we search for the meaning of the words is to imagine what the speaker´s communicative intention might be" [p. 27] which refers to the pragmatics of the spoken words. As was already alluded to earlier, current dialogue and agent systems are prone for misunderstandings and failed comprehension attempts. Although the reasons for this are manifold, an important explanation is the fact that basic needs and customs of the human users are neglected.

4.5 Implications

The obvious consequence of these considerations is thus to try to implement theory-of-mind-like abilities. This includes that the agent has to be "aware" of his own abilities and knowledge about the human interaction partner. Therefore a user model is needed which incorporates global knowledge on human needs and states. Containing basic knowledge on human abilities, knowledge, states, etc. and interaction abilities this ToM module enables the agent to verify the knowledge, beliefs, emotions, etc. of the user and to progressively build common ground with the user [Krämer, 2008].

Krämer [2008] summarizes that "it might be stated that the different models on common ground, perspective taking, imputing one´s knowledge and ToM show major similarities with regard to the fact that all propose that humans possess a direct but implicit knowledge on other humans (…) that form a starting point for mutual comprehension. Building on this, the dialog can be used to clarify and broaden mutual knowledge by means of grounding processes." Most current agent systems lack both: a theory of its own mind and a complete user module that can be compared to a ToM. Krämer [2008] discusses a couple of approaches that implemented a ToM module in an agent [e.g. Marsella and Pynadath, 2005; Traum, 1996; Breazeal et al. 2004] and concludes that they are promising but will have to be broadened. To successfully implement a ToM module there it needs more than simply implementing rules or knowledge: "The bottom line of the idea of mentalising is that we predict what other individuals will do in a given situation from their desires, their knowledge and their beliefs, and not from the actual state of the world" [Frith and Frith, 2003, p. 6].

Within the state of the art of artificial intelligence it is a long way towards a successful development of a ToM module, but some principles promise efficient mechanisms to avoid misunderstandings in human-agent/robot-interaction. Also, the possibility of the system to give feedback should be fostered. Krämer [2005] observed that people repeated their utterances when the agent did not give feedback in an appropriate time insinuating the agent did not understand what they said. Thus the statement: "One cannot not communicate." [Watzlawick et al., 1967] holds also for agents. Immediate feedback (also about the delay of appropriate feedback) is crucial to avoid misunderstanding and frustration.

5 Conclusion

In sum it can be stated that we have to carefully consider basic human abilities that we take for granted in everyday communication in order to be able to built artificial entities that are able to not only engage in sensible short time conversations but also develop relationships. With regard to the former, perspective taking and ToM will be crucial in order for the robot/agent to predict what effect a specific utterance will have on the user – given that the user will construct the meaning against the background of his/her human abilities. With regard to developing relationships, additionally a possibility to build common ground has to be established as the robot/agent has to have the ability to build on joint experiences and former dialogues.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Program [FP7/2007-2013] under grant agreement no. 231868.

References

[Baron-Cohen, 1995] S. Baron-Cohen. *Mindblindness. An essay on autism and theory of mind*. MIT Press, Cambridge, 1995.

[Baumeister and Leary, 1995] R.F. Baumeister and M.R. Leary. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(497-529), 1995.

[Berscheid, 1985] E. Berscheid. Interpersonal attraction. In G. Linzey and E. Aronson, eds. *Handbook of Social Psychology* (Vol. 2, pp. 413-484). Random House, New York, 1985.

[Berscheid and Reis, 1998] E. Berscheid and H.T. Reis. Attraction and close relationships. In D.T. Gilbert, S.T. Fiske and G. Lindzey, eds. *The handbook of social psychology* (4th ed., Vol. 2, pp. 193-281). McGraw-Hill, New York, 1998.

[Berscheid and Walster, 1978] E. Berscheid and E. Walster. *Interpersonal attraction*. Addison-Wesley, Reading, MA, 1978.

[Breazeal et al., 2004] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby and B. Blumberg. Learning from and about Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots. In L. Rocha, F. Almedia e Costa, eds. *Artificial Life (*pp. 31-62). MIT Press, Cambridge, MA, 2004.

[Clark, 1992] H.H. Clark. *Arenas of language use*. University of Chicago Press, Chicago, 1992.

[Clark and Carlson, 1981] H.H. Clark and T.B. Carlson. Context for comprehension. In J. Long and A. Baddeley, eds., *Attention and performance IX*. (pp. 313-330). Erlbaum, Hillsdale, NJ, 1981.

[Curtis and Miller, 1986] R.C. Curtis and K. Miller. Believing another likes or dislikes you: behaviours making the beliefs come true. *Journal of Personality and Social Psychology*, 51(284-290), 1986.

[Dennett, 1987] D.C. Dennett. *The intentional stance*. MIT Press, Cambridge, 1987.

[Dion, Berscheid and Walster, 1972] K. Dion, E. Berscheid and E. Walster. What is beautiful is good. *Journal of Personality and Social Psychology*, 24(285-290), 1972.

[Duck and Pittman, 1994] S.W. Duck and G. Pittman. Social and personal relationships. In M.L. Knapp and G.R. Miller, eds, *Handbook of interpersonal*

communication (2nd ed., pp. 676-695). Sage, Thousand Oaks, CA, 1994.

[Frith and Frith, 2003] U. Frith and C.D. Frith. Development of neurophysiology of mentalizing. *Phil. Trans. R. Soc. Lond B Biol Sci*, 358 (459-473), 2003.

[Gold et al., 1984] J.A. Gold, R.M. Ryckman and N.R. Mosley. Romantic mood induction and attraction to a dissimilar other: Is live blind? *Personality and Social Psychology Bulletin*, 10(358-368), 1984.

[Hatfield et al., 1978] E. Hatfield, G.W. Walster and E. Berscheid. Equity: Theory and research. Allyn & Bacon, Boston.

[Homans, 1961] G.C. Homans. *Social behaviour: Its elementary forms*. Harcourt Brace, New York, 1961.

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

[Kelley and Thibaut, 1978] H.H. Kelley and J. Thibaut. *Interpersonal relations: A theory of interdependence*. Wiley, New York, 1978.

[Krämer, 2008] N.C. Krämer. Theory of Mind as a theoretical prerequisite to model communication with virtual humans. In I. Wachsmuth and G. Knoblich, eds. *Modeling communication with robots and virtual humans* (pp. 222-240). Springer, Berlin, 2008.

[Krämer, 2005] N.C. Krämer. Social communicative effects of a virtual program guide. In T. Panayiotopoulos et al.,eds. *Intelligent Virtual Agents 2005*, pp. 442-543. Springer, Hamburg, 2005.

[Krauss and Fussell, 1991] R.M. Krauss and S.R. Fussell. Perspective taking in communication: Representation of others´ knowledge in reference. *Social Cognition*, 9(2-24), 1991.

[Kubitschek and Hallinan, 1998] W.N. Kubitschek and M.T. Hallinan. Tracking and sutdents' friendships. *Social Psychology Quarterly*, 61(1-15), 1998.

[McPherson et al., 2001] M. McPherson, L. Smith-Lovin and J.M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(415-444), 2001.

[Marsella et al., 2005] S.C. Marsella, D.V. Pynadath. Modeling influence and theory of mind. Artificial Intelligence and the Simulation of Behavior, Joint Symposium on Virtual Social Agents, 199-206, 2005.

[Nickerson, 1999] R.S. Nickerson. How we know – and sometimes misjudge – what others know: Imputing one´s knowledge to others. *Psychological Bulletin*, 125(737-759), 1999.

[Norman, 1988] D.A. Norman. *The design of everyday things*. Doubleday, New York, 1988.

[Premack and Premack, 1995]. D. Premack and A.J. Premack. Origins of human social competence. In M.S. Gazzaniga, ed. *The cognitive neurosciences* (pp. 205-218). MIT Press, Cambridge, MA, 1995.

[Premack and Woodruff, 1978] D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, 4(512-526), 1978.

[Rusbult, 1983] C.E. Rusbult. A longitudinal test of the investment model: The development (and deterioration) of satisfaction and commitment in heterosexual involvement. *Journal of Personality and Social Psychology*, 45(101-117) 1983.

[Swann et al., 1992] W.B. Swann, A. Stein-Seroussiand and S.E. McNulty. Outcasts in a white-lie society: the enigmatic worlds of people with negative self-concepts. *Journal of Personality and Social psychology*, 62(618-624), 1992.

[Swap, 1977] W.C. Swap. Interpersonal attraction and repeated exposure to rewarders and punishers. *Personality and Social Psychology Bulletin*, 3(248-251), 1977.

[Thibaut and Kelley, 1959] J.W. Thibaut and H.H. Kelley. *The social psychology of groups*. Wiley, New York, 1959.

[Toby and Cosmides, 1995] J. Toby and L. Cosmides. Foreword. In S. Baron-Cohen, ed. *Mindblindness. An essay on autism and theory of mind*. MIT Press, Cambridge, 1995.

[Traum, 1996] D. Traum. Conversational Agency: The Trains-93 Dialogue Manager. In *Proceedings of the Twente Workshop on Language Technology 11: Dialogue Management in Natural Language Systems*, 1-11, 1996.

[Van Vugt et al., 2006] H.C. Van Vugt, E.A. Konijn, J. F. Hoornand and J. Veldhuis. Why Fat Interface Characters Are Better e-Health Advisors. In J. Gratch et al., eds. *IVA 2006*, LNAI 4133, pp. 1-13. Berlin: Springer, 2006.

[Walster et al., 1996] E. Walster, Aronson, D. Abrahams, L. Rottman. Importance of physical attractiveness in dating behavior. *Journal of Personality and Social Psychology*, 5(508-516), 1966.

[Watzlawick, et al., 1967] P. Watzlawick, J.H. Beavin and D.D. Jackson. Pragmatics of human communication. A study of interactional patterns, pathologies, and paradoxes. W.W. Norton & Co, NY, 1967.

[Whiten, 1991] A. Whiten. *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. Basil Blackwell, Oxford, 1991.

[Zajonc et al., 1989] R.B. Zajonc, S.T. Murphy and M. Inglehart. Feeling and facial efference: Implication of the vascular theory of emotion. *Psychological Review*, 96(395-416), 1989.

Methodological Considerations for Long-Term Experience with Robots and Agents

Tina Ganster, Sabrina C. Eimler, Astrid M. von der Pütten, Laura Hoffmann, Nicole C. Krämer

University of Duisburg-Essen
Forsthausweg 2
47048 Duisburg
Germany

tina.ganster@stud.uni-due.de, {sabrina.eimler, astrid.von-der-puetten, laura.hoffmann, nicole.kraemer}@uni-due.de

Abstract

This paper concentrates on the integration of methodological approaches, addressing the complexity and potential problems of research on long-term relationships with robots and agents. By presenting different categories of methods and outlining their potential and their benefits with regard to long-term human-agent interactions, solutions are suggested.

1. Introduction

Artificial entities have always been a popular part of science fiction stories. An often recurring aspect of these stories is that those characters are interactive entities that answer to their human conversational partners in an almost natural way and as a consequence, people get attached to them (e.g. Number Five or R2D2 from Star Wars). At the present time, we are still far away from this fictional vision but research already contributes to a similar development by promoting the creation and exploration of interactive agents and robots present in our lives. This includes continuous evaluation and improvement of their interaction with human beings. To achieve this, besides evaluating the agents' performance in terms of usability [e.g. Ruttkay and Pelachaud, 2004], a special focus has to be laid on users' personal perception of a system. Aspects like the systems' helpfulness, believability or the users' satisfaction with and trust in the agent as well as the perceived engagement are important prerequisites for improving system designs over time. In the early days of agent research, researchers concentrated on short-term human-agent interaction. With steady progress in research, the focus more and more shifts towards the development of systems that are not merely temporarily applied within a laboratory environment but rather enable (and encourage) long-term interactions within a user's natural environment. Research on long-term relationships regularly turns out to be costly and time-consuming since extensive preparation and repeated use of expensive measurement techniques have to be considered. Another crucial point also is the involvement of people who participate in an experiment for a longer time and thus have to be paid and taken care of (because of possible moral and ethical concerns). Since research in this area is only just emerging, a lack of expertise cannot be denied. Often, methods and instruments that have proven to be helpful in short-term studies are transferred without knowing whether these are applicable and adequate for long-term interactions.

Against this background we intend to draw attention to several methodological aspects and discuss whether these may be applicable in long-term interactions. With the objective of improving data quality, reducing unnecessary costs and efforts, and increasing comparability of results throughout this research area, we suggest several aspects in need of improvement to motivate and facilitate research particularly on long-term interactions and with regard to an agent's social behavior.

The first aspect we consider to be important is the rather inconsiderate usage of objective and subjective measurement methods. We provide a list of widely established instruments and their respective benefits to the investigation of long-term relationships with agents. The second aspect particularly deals with the need to arrive at and to make use of more standardized measurement instruments. Within the field of objective measurement methods, we point out the possibility to use video and audio analysis to determine relevant changes in the course of the temporal development in the user-agents relationship.

2. Objective vs. Subjective Measurement Methods

According to Krämer [2002], relevant aspects of human-agent interaction can be categorized by distinguishing objectively measureable and subjectively measureable

ones. The former comprise for instance social effects and the user's experience during the interaction (often assessed by means of self-report scales) while aspects like user behavior and efficiency can be measured objectively (through observation).

Each way of measuring covers different phenomena in human-agent interaction but because of their specific strengths and weaknesses, they can be effectively combined to make results more conclusive. With regard to long-term interactions, particularly the usage of subjective measurement methods has to be reconsidered and adapted to serve this special purpose. In the following we will review subjective and objective measurements with regard to their respective benefits and shortcomings in long-term experiments.

Instruments of subjective measurement are commonly used in psychological research. A major part of it can be summarized as surveys [Fife-Schaw, 2000] that are usually filled in by participants at one or several points in time during or at the end of the study. With respect to human-agent interaction it can be stated that surveys are not conducted during the actual interaction but usually afterwards. Few measurement scales are particularly (and widely) used for the evaluation of human-agent interaction as for example the Agents Persona Instrument (API) by Baylor and Ryu [2003] and the Attitude Towards Agents Scale (ATAS) [Van Eck and Adcock, 2003]. With regard to the acceptance of agents, several approaches using survey are conceivable, including a general evaluation (using questionnaires with open- or closed-ended questions), evaluation of appearance [e.g. card sort assignments; Cowell and Stanney, 2003], perceived efficiency [e.g. Krämer and Nitschke, 2002], believability and trust in a system [e.g. Rickenberg and Reeves, 2000; Sproull et al., 1996; Krämer et al., 2005] as well as the personal feelings associated with the interaction [e.g. Krämer and Nitschke, 2002]. The measurement of social effects may comprise evaluations of socio-emotional aspects and person perception [e.g. Krämer, 2001] as well as para-social interaction [cf. Hartmann et al., 2001].

Surveys used in psychological research can be classified by their respective mode of data collection [Biemer and Lyberg, 2003]. Each mode may offer benefits as well as weaknesses, depending on the kind of data being gathered. Face-to-face interviews may offer good data quality because they provide high flexibility. But interviewer effects and social desirability have to be taken into account [Biemer and Lyberg, 2003]. Hoffmann et al. [2009] for instance showed that interviewer effects also apply to an agent whose performance was evaluated better when participants were confronted with the agent in contrast to filling out a paper-pencil-questionnaire. This also shows that the tendency for interviewer effects and socially desirable answers can be decreased through electronic or paper questionnaires.

Additionally, with respect to long-term interactions, it has to be pointed out that repeated surveys are much cheaper and easier to integrate when not performed face-to-face. In line with this, keeping diaries may as well turn out to be particularly helpful in assessing interaction placed in the user's natural environment and at the same time provide qualitatively valuable data.

In sum, subjective measurement methods provide the opportunity to capture personal and subjective aspects of human-agent interaction. This kind of evaluation is essential for the aim of improving an agent's social behavior and thus its acceptance and likeability. As illustrated, subjective methods are likely to be of low complexity and more economic in contrast to objective methods (as will be shown later). But possible deficits have to be taken into account, for example interviewer effects, response rates or the fact that a survey normally is taken after the actual interaction (so the response is delayed) and only captures a snapshot in time. The latter aspect particularly represents a problem in long-term interactions since evaluation and memory performance are found to be distorted by primacy and recency effects when referring to a longer period of time. Another potential problem is referred to as reconstructive memory, meaning that the memories about an event are distorted due to incompatible experiences after the event [Atkinson & Shiffrin, 1968]. In the case of long-term interactions this implies that the emotions at the end of an experiment may dominate and overwrite the memory about emotions occurring at the beginning.

Objectively measureable effects of human-agent interaction relate to performance and user behavior from which an implicit measure of acceptance and likeability can be derived. A behavior that may directly refer to acceptance is for example the decision to use an agent for specific tasks or not [Krämer, 2002]. This aspect is particularly relevant in long-term studies where the participant is free to decide. Activation or arousal can also be regarded as indicators for the dimension of acceptance. According to Krämer [2002], the efficiency of a system can be measured through learning effects, task accomplishments or the change of performance through social inhibition [cf. Rickenberg and Reeves, 2000]. Another dimension of objectively measureable aspects covers all kinds of user behavior, such as the use of natural language [cf. Krämer and Nitschke, 2002], impression-management behavior [measurable via social desirable behavior in questionnaire items; e. g. Sproull et al., 1996], reciprocal helping [behavior in social dilemmas; cf. Nass and Moon, 2000], attention, physiological arousal, or body movement.

A widely used instrument for objective measurements in research on human-computer and human-agent interaction is *eye tracking*. This term refers to a number of different techniques to conduct research on eye or head movement, respectively, that vary with regard to precision and intrusiveness [Duchowski, 2003]. Since intrusiveness may cause discomfort to the wearer and thus lead to distorted results, modern eye tracking technology tends to become more unintrusive, leaving out head fixation to allow a more natural user behavior. With the help of eye tracking, researchers can determine users' (focus of) attention (e.g. by observing gaze direction) and aspects perceived most interesting or cognitively challeng-

ing (e.g. by measuring duration of eye-gaze behavior). With regard to long-term interaction, it can be pointed out that unintrusive methods of eye tracking may be especially relevant to research but costs of excessive or even permanent usage have to be taken into account. Regular short-term use within a long-term experiment may prove as feasible.

As a second method, *psychophysiology* shall be outlined here. The term refers to a number of instruments measuring physiological arousal [Graham & Schandry, 2009] that are also commonly used for research on human-agent interaction. Among the techniques that are used most often today is Electrodermal Response (EDR), which cannot identify the quality of an emotion but is nevertheless highly sensitive. Furthermore, Electrocardiography (ECG) as well as Electroencephalography (EEG) are non-invasive methods, working with skin electrodes attached to the human body. Since experiments on long-term interactions tend to take place in the user's natural environment, only mobile instruments (e.g. ECG) are efficient in this respect.

One of the most complex and technically advanced instruments for objective measurements is the *Functional Magnetic Resonance Imaging (fMRI)*, a rather recently developed method of neuroimaging [Ogawa et al., 1990] that is able to visualize the level of activity of different parts of the human brain. FMRI-based experiments normally consist of a sequence of different single scans. Although the method is non-invasive, the technique is highly sensitive to movement so the respective body part has to be fixated, extremely decreasing the naturalness of the interaction situation. For this reason, fMRI is probably not functionally adequate for research on long-term interactions.

Another series of methods classified as objective is the *analysis of audio and video material* recorded during the interaction [e.g. Gratch et al, 2006, 2007; Kang et al., 2008]. Audio material can be analyzed to identify certain characteristics of the subject's use of language. Parameters for language analysis may for example be the number or lengths of utterances, hesitations, pauses or the number of incomplete words or conversational fillers [e.g. Gratch et al., 2006]. To increase reliability, the quantitative analysis of natural language can and should be combined with qualitative analysis, such as analysis on Self-Disclosure [Moon, 2000; von der Pütten et al., submitted] and Conversation Analysis [Ten Have, 2007].

Video analysis can be classified as a less intrusive and more accurate variation of direct observation since the subject is not observed during conversation, but subsequently. This also enables the observer to repeatedly look into the material thereby increasing reliability of the assessment. In video analysis, particularly nonverbal behavior is of interest. Since nonverbal communication is a complex phenomenon, one has to carefully select appropriate methods capable of capturing all relevant aspects [Krämer, 2008a]. On the one hand, the users' general behavior can be observed with respect to their way of interacting with an agent which can for instance be supported by systems like the Bernese System for Time Series Notation [Frey et al., 1980]. On the other hand, the focus can be laid on facial expressions. One of the oldest methods for the measurement and description of facial activity, the Facial Action Coding System (FACS) proposed by Ekman and Friesen [1978] is still used today. However, in the majority of studies, these dynamic aspects of long-term human-agent interaction are not taken into account, (see below).

All of these aspects are objectively observable, but immediate insights into the subjective user experience are not possible. Yet, certain insights can be gained from these observations, for example that a user would only show social behavior (e.g. impression management) to something he/she perceives as a social being (for an overview about social effects of artificial entities see [Krämer, 2008b]).

A large number of objective measurement instruments were presented, including the (potential) shortcomings and strengths for their application in long-term experiments. In contrast to subjective measurements, they allow researchers to capture user behavior and derive emotional reactions along a temporal dimension which represents an important aspect of long-term interactions. When considering a methodological approach, the different objective methods have to be weighted with regard to the respective studies' objective. Some instruments relating to physiological arousal do not allow the participant to move freely (e.g. EDA). This may result in discomfort for the participant who may not be able to act naturally on the one hand; on the other hand, this may interfere with other objectives, e.g. the observation of nonverbal behavior. In general, the complexity of most objective methods represents an obstacle in long-term interactions with respect to natural environments. Here, video and audio analyses seem to be quite unproblematic and unintrusive instruments. With regard to the effects that are observable via objective measurements, it can be stated that they can only indirectly capture the quality of internal emotional states. The level of arousal does not always give an indication of the quality of the corresponding emotion. Similar problems occur within natural language and video analysis. There exist certain methods to identify and quantify the observed behavior but the tendency for interpretation cannot entirely be ruled out.

A rather advantageous aspect of objective research methods that has up to now not captured many researchers' attention is the possibility to make use of the subject's output data (e.g. eye gaze behavior or heart rate response). Since objective measures are taken during interaction, they can be used as inputs in the further course of the interaction. This can be applied to various aspects of human-agent interaction: In human-human communication turn-taking is an important mechanism in the course of which a number of behavioral indicators for holding, taking and giving the turn can be observed. As eye contact is crucial in this regard, eye tracking data might be used as input information for the agent, either to know when he is to speak or in order to signal the user

that he has the floor. This has for example been implemented in the Real Estate Agent that uses eye contact to realize turn-taking [Bickmore, 2003; Cassell et al., 2002] as well as Jonsdottir and Thórisson's [2009] system that analyses pauses in speech flow to regulate turn-taking. With respect to showing interest in a user, eye-gaze behavior can be used to establish mutual eye contact between the agent and the user [Bee et al., 2009] on the one hand. On the other hand, the agent might make use of this information in order to change its behavior and make the user feel more comfortable.

In this respect, objective methods have benefits as well as disadvantages compared to subjective ones. Shortcomings and benefits of each may be outweighed by applying an appropriate mixture of both kinds to cover more relevant aspects of human-agent interaction. A common example would be using psycho physiological instruments to determine the level of arousal while capturing the quality of the corresponding emotion via questionnaire.

3. Using Standardized Instruments

Benefits resulting from the usage of applicable and appropriate methods are dependent on their validity. With content validity referring to the best possible operationalization of a concept, it is directly related to the topic of standardized subjective instruments that have been tested with regard to their validity and reliability Since the area of human-agent interaction is a rather new topic in psychological research, only few of those exist. Some of them were already mentioned at the beginning, such as the Agents Persona Instrument (API) by Baylor and Ryu [2003] and the Attitude Towards Agents Scale (ATAS) [Van Eck and Adcock, 2003; Beun et al., 2003]. Bartneck et al. [2009] stress the need to arrive at more standardized measurement instruments for the categories they have identified as key concepts in human-robot interaction, such as likeability, perceived intelligence, and perceived safety. They propose a set of semantic differential questionnaires, the *Godspeed questionnaires*, covering these categories.

The use of psychological measurement instruments was already suggested, since in human-agent interaction, person perception may play an important role. When it comes to measuring this perception, one can observe a tendency to use ad-hoc questionnaires including single items representing complex dimensions of person perception. Such ad-hoc methods do not only increase difficulties in comparing and contrasting results from different studies, one also has to question the validity of such a questionnaire in contrast to an established and widely used psychological scale. Scales that may be useful in human-agent interaction are for example the Positive and Negative Affect Scale (PANAS) [Watson et al., 1988], which is often used when emotional experiences are evaluated, as well as the Self Assessment Mannequin Scale (SAM) [Lang et al., 1993]. There also exist transferrable questionnaires on person perception [Hurwitz et al., 1975] and the concept of presence (e.g. the Temple Presence Inventory by Lombard [2005]). Although some standards with respect to objective measurements have been listed, it has to be stated that there is a larger amount of standardized subjective measurements compared to objective ones. Creating objective measurement standards may turn out to be difficult since various different systems can be used to record data. These systems determine which aspects can be analyzed and whether standards can be applied.

4. Process Data

Looking at human-agent relations, we always find the interaction to be a process of a certain duration including a certain number of incidents and actions. As already pointed out, subjective measurements only allow the caption of the experience afterwards and at a specific point in time (except for the diary method). Changes that occurred during the respective period of time are not taken into account as was outlined in section **2** already.

Objective measures are more appropriate when relevant changes should be determined. For instance, Wada and Shibata [2006] repeatedly collected urine samples in the context of their long-term study with seal robots used for therapy in a care house. When presenting different methods of objective measuring, it became clear that a large part of these instruments allow the effects to be measured along a timeline, e.g. psychophysiological measures, and to relate possible changes to the single components of the interaction. The same opportunity is also offered by analyzing natural language and video material. But up to now, this has scarcely been made use of. Despite researchers stressing the importance of methods enabling a dynamic, procedural mapping of nonverbal behavior [Cappella & Palmer, 1990; Monge & Kalman, 1996], most studies conducted rely on "distributional" instead of "temporal" data [Cappella & Palmer, 1990]. One of the few measuring systems for human movements that incorporate temporal aspects is the Bernese System for Time Series Notation [Frey et al., 1983; for an overview see Donaghy, 1989]. Within the research area of human-agent interaction, the opportunity to make use of "temporal" data with regard to nonverbal and verbal behavior is still widely missing.

5. Conclusion

With respect to long-term human-agent interaction, we stressed the need for the use of applicable research methods capable of handling the complexity of long-term interactions. We suggested possibilities to overcome shortcomings of previous research by presenting different

categories of research methods and outlining their benefits with regard to human-agent relationship research.

When it comes to deciding whether to use objective or subjective measurement instruments, one has to consider not only the objective of the study but also the various benefits and disadvantages of methods presented. With regard to objective measurement methods it was shown that many of them still hold hidden potentials advantages to make evaluation more complex and reliable. We suggested that it would be advantageous to combine both kinds of methods to increase data quality and reliability.

The disadvantages of non-standardized, ad-hoc questionnaires were pointed out resulting in the suggestion to make use of already validated psychological scales and those standards existent among objective methods.

However, joint efforts have still to be undertaken to improve all kinds of methods. Advancements in the research area will depend on researchers´ willingness to agree on (standardized) measures and to employ process measures even if this means additional effort.

References

[Atkinson and Shiffrin, 1968] R. C. Atkinson and R. M. Shiffrin. Human Memory: A Proposed System and Its Control Processes. In K.W. Spence & J.T. Spence (Eds.). *The Psychology of Learning and Motivation*. Vol 2. Acad. Press, New York, 1968.

[Bartneck et al., 2009] C. Bartneck, E. Croft, D. Kulic, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), pp. 71-81, 2009.

[Baylor and Ryu, 2003] A. L. Baylor and J. Ryu. *The API (Agent Persona Instrument) for Assessing Pedagogical Agent Persona*. Honolulu, Hawaii: ED-Media, 2003.

[Bickmore, 2003] T. W. Bickmore. *Relational Agents: Effecting Change through Human-Computer Relationships*. PhD, MIT, 2003.

[Biemer and Lyberg, 2003] P. Biemer and L. Lyberg. *Introduction to Survey Quality*. John Wiley & Sons, New Jersey, 2003.

[Bee et al., 2009] N. Bee, E. André and S. Tober. Breaking the Ice in Human-Agent Communication: Eye-Gaze Based Initiation of Contact with an Embodied Conversational Agent Export. *9th International Conference on Intelligent Virtual Agents (IVA 2009)*, pp. 229-242. Springer, Berlin, 2003.

[Capella and Palmer, 1990] J. N. Cappella and M. T. Palmer. The structure and organization of verbal and non-verbal behavior: Data for models of production. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology*, pp. 141-161. John Wiley and sons, Chichester, 1990.

[Cassell et al., 2002] J. Cassell, T. Stocky, T. Bickmore, Y. Gao, Y. Nakano, K. Ryokai, D. Tversky, C. Vaucelle and H. Vilhjálmsson. MACK: Media lab Autonomous Conversational Kiosk. In *Proceedings of Imagina 2002*. February 12-15 2002, Monte Carlo.

[Cowell and Stanney, 2003] A. J. Cowell, and K. M. Stanney. Embodiment and interaction guidelines for designing credible, trustworthy embodied conversational agents. In T. Rist, R. Aylett, D. Ballin, and J. Rickel (Eds.), *Proceedings of the Fourth International Working Conference on Intelligent Virtual Agents (IVA2003)*, pp. 301-309. Springer Verlag, Berlin, 2003.

[Donaghy, 1989] W. C. Donaghy. Nonverbal communication measurement. In P. Emmert and L. Barker (Eds.), *Measurement of communication behaviour*, pp. 296-332. White Plains, NY: Longman, 1989.

[Duchowski, 2003] A. T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, London, 2003.

[Ekman and Friesen, 1978] P. Ekman and W. Friesen. *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, California, 1978.

[Fife-Shaw, 2000] C. Fife-Schaw. Surveys and Sampling Issues. In G. M. Breakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology*, pp. 88-104. Sage Publications, London, 2000.

[Frey et al., 1983] S. Frey, H.-P. Hirsbrunner, A. Florin, W. Daw, and R. Crawford. A unified approach to the investigation of nonverbal and verbal behavior in communication ressearch. In W. Doise and S. Moscovici (Eds.), *Current issues in European Social Psychology*, pp. 143-199. Cambridge University Press, Cambrigde, 1983.

[Frey et al., 1981] S. Frey, H.-P. Hirsbrunner, J. Pool and W. Daw. Das Berner System zur Untersuchung nonverbaler Interaktion: I. Die Erhebung des Rohdatenprotokolls. In P. Winkler (Ed.). *Methoden der Analyse von Face-to-Face-Situationen*, pp. 203-236. Metzler, Stuttgart, 1981.

[Graham and Schandry, 2009] K. Graham and R. Schandry. *Lehrbuch Psychophysiologie: Körperliche Indikatoren psychischen Geschehens.* Beltz Psychologie Verlags Union, Weinheim, 2009.

[Gratch et al., 2006] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales and R. J. van der Werf. *Virtual Rapport*. Marina del Rey, CA, 2006.

[Gratch et al., 2007] J. Gratch, N. Wang, J. Gerten, E. Fast and R. Duffy. Creating rapport with virtual agents. *7th International Conference on Intelligent Virtual Agents*. Paris, France, 2007.

[Hartmann et al., 2001) T. Hartman, C. Klimmt and P. Vorderer. Avatare: Parasoziale Beziehungen zu virtuellen Akteuren. *Medien- und Kommunikationwissenschaft*, 49, pp. 350-268, 2001.

[Hoffmann et al., 2009] L. Hoffmann, N. C. Krämer, A. Lam-Chi and S. Kopp. Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In Z. Ruttkay (Eds.), *IVA 2009*. LNAI 5773, pp. 159-165. Springer-Verlag, Berlin, 2009.

[Hurwitz, 1975] D. Hurwitz, N. Wiggins, and L. Jones. A semantic differential for facial attribution: The face differential. Bulletin of the Psychonomic Society (6), pp. 370-372, 1975.

[Jonsdottir and Thórisson, 2009] G. R. Jonsdottir and K. R. Thórisson. Teaching Computers to Conduct Spoken Interviews: Breaking the Realtime Barrier with Learning. In Z. Ruttkay (Ed.), *IVA 2009*. LNAI 5773, pp. 446-459. Springer-Verlag, Berlin, 2009.

[Kang et al, 2008] S.-H. Kang, J. Gratch, N., Wang and J. H. Watts. Does Contingency of Agents' Nonverbal Feedback Affect Users' Social Anxiety? *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pp. 120-127. Estoril, Portugal: International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[Krämer, 2001] N. C. Krämer. *Bewegende Bewegung. Sozio- emotionale Wirkungen nonverbalen Verhaltens und deren experimentelle Untersuchung mittels Computeranimation*. Pabst, Lengerich, 2001.

[Krämer, 2002] N. C. Krämer. Können virtuelle Helfer uns wirklich helfen? – Methoden zur Evaluation von anthropomorphen Mensch-Maschine-Schnittstellen. *Zeitschrift für Medienpsychologie*, 14 (1), pp. 34-37, 2002.

[Krämer, 2008a] N. C. Krämer. Nonverbal Communication. In J. Blascovich and C. Hartel (Eds.), *Human behavior in military contexts*, pp. 150-188. The National Academies Press, Washington, 2008.

[Krämer, 2008b] N. C. Krämer. *Soziale Wirkungen virtueller Helfer. Gestaltung und Evaluation von Mensch- Computer-Interaktionen*. Kohlhammer, Stuttgart, 2008.

[Krämer et al., 2005] N. C. Krämer, I. Iurgel and G. Bente. Emotion und motivation in embodied conversational agents. In L. Canamero (Ed.), *Proceedings of the Symposium "Agents that Want and Like" (AISB 2005)*, pp. 55-61. SSAISB, Hatfield, 2005.

[Krämer and Nitschke, 2002] N. C. Krämer and J. Nitschke. Ausgabemodalitäten im Vergleich: Verändern sie das Eingabeverhalten der Benutzer? In R. Marzi, V. Karavezyris, H.-H. Erbe and K.-P. Timpe (Eds.), *Bedienen und Verstehen. 4. Berliner Werkstatt Mensch-Maschine-Systeme*, pp. 231-248. VDI-Verlag, Düsseldorf, 2002.

[Lang et al., 1993] P. J. Lang, M. Greenwald, M. M. Bradley and A. O. Hamm. Looking at pictures: Evaluative, facial, visceral, and behavioral responses. Psychophysiology, 30, pp. 261-273, 1993.

[Lombard, 2005] M. Lombard, M. Temple Presence Inventory (TPI) questionnaire items and factor structures. Retrieved November 26, 2009 from http://astro.temple.edu/~lombard/research/P2scales_11-04.doc

[Monge and Kalman, 1996] P. R. Monge and M. E. Kalman. Sequentiality, simultaneity and synchronicity in human communication. In J. H. Watt and C. A. VanLear (Eds.), *Dynamic patterns in communication processes*, pp. 71-92. Sage Publications, Thousand Oaks, 1996.

[Moon, 2000] Y. Moon. Intimate Exchanges: Using Computers to Elicit Self-disclosure from Consumers. *Journal of Consumer Research*, 26 (4), pp. 324-340, 2000.

[Nass and Moon, 2000] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56 (1), pp. 81-103, 2000.

[Ogawa et al., 1990] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. In *Proceedings of the National Academy of Sciences of the United States of America*, 87 (24), pp. 9868-9872, 1990.

[Rickenberg and Reeves, 2000] R. Rickenberg and B. Reeves. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *Letters of CHI 2000*, pp. 49-56, April 2000.

[Ruttkay and Pelachaud, 2004] Z. Ruttkay and C. Pelachaud (Eds.). *From Brows to Trust: Evaluating Embodied Conversational Agents* (Human Computer Interaction Series, Vol. 7). Kluwer, New York, 2004.

[Sproull et al., 1996] L. Sproull, M. Subramani, S. Kiesler, J. H. Walker and K. Waters. When the interface is a face. *Human Computer Interaction*, 11 (2), pp. 97-124, 1996.

[Ten Have, 2007] P. Ten Have. *Doing Conversation Analysis* (2nd ed.). London: Sage, 2007.

[Van Eck and Adcock, 2003] R. Van Eck and A. Adcock. Reliability and factor structure of the Attitude Towards Agent Scale (ATAS). *Paper presented at the AERA*, Chicago, IL, 2003.

[Von der Pütten et al., submitted] A. M. Von der Pütten, N. C. Krämer, J. Gratch, and S. H. Kang. "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Manuscript submitted for publication.*

[Wada and Shibata, 2006] K. Wada, T. Shibata. Robot Therapy in a Care House - Its Sociopsychological and Physiological Effects on the Residents. *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, 2006.

[Watson et al., 1998] D. Watson, L. A. Clark and A. Tellegen. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070, 1998.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 231868.

Acceptance and Use of a Zoomorphic Robot in a Domestic Setting

Tineke Klamer, Somaya Ben Allouch
University of Twente
Faculty of Behavioural Sciences
Department Media, Society and Organizations
Drienerlolaan 5, 7522 NB Enschede, the Netherlands
{T.Klamer, S.BenAllouch}[at]utwente.nl

Abstract

The study presented in this article aims to improve our understanding of how people use zoomorphic robots in domestic environments in general and, in particular, whether people are able to build (long-term) relationships with these robots. The influences of social and hedonic factors were studied, in addition to the utilitarian factors of the Technology Acceptance Model (TAM). Three participants interacted with the Nabaztag, a zoomorphic robot, for 10 days in their own home environment. No evidence was found that hedonic factors were important for the acceptance of the Nabaztag. However, hedonic factors did seem to be important for building a relationship with the Nabaztag. Social factors did seem to be important for the acceptance of robots, but they did not seem to be important for building a relationship with the Nabaztag. Last, the results showed a relationship between naming the Nabaztag and building a relationship with it.

1 Introduction

Imagine the year 2019. Mr. Smith, 90 years old, is still able to live autonomously thanks to his social robot Suzy. Suzy cleans his house, washes his clothes, gets his groceries, cooks his meals everyday, does his dishes, plays card games with him, discusses politics and monitors his health. Mr. Smith states: "She is my best friend and I can not live without her any more." Will this be reality ten years from now?

It is assumed that in the near future, social robots will be able to aid the elderly to live longer autonomously in their own homes. In the near future, robots will be able to, for example, do household tasks for them, monitor their health and be a social companion. Therefore it is important to study the acceptance and use of social robots, so that future social robots can be adapted to the wishes and demands of the elderly, which is important for the future diffusion and adoption of robotic technology.

All definitions of a social robot are built upon the same idea. That is: social robots are robots that interact via human social rules, e.g. [Looije, Chossen and Neerincx, 2006] [Leite et al., 2008] [Dautenhahn, 2002]. Social robots appear in different forms. There are for example, humanoid, mechanoid and zoomorphic robots. Humanoid robots are social robots that resemble human beings. Second, mechanoid robots are robots that are more mechanical looking and are more often used in practical situations (e.g. to rescue people, to help them in factories). Finally, zoomorphic robots are robots that resemble animals, e.g. AIBO, Sony's dog-like robot. In this study we specifically focus on zoomorphic robots.

1.1 Acceptance and Use of Social Robots

Acceptance of robots is assumed to be different from acceptance of other technical innovations. The original Technology Acceptance Model (TAM) aims to understand the utilitarian, productivity oriented, use of technology [Davis, Bagozzi, & Warshaw, 1992]. But besides utilitarian use of technology, there is also a hedonic, pleasure oriented use of technology [Heijden, 2004]. For example, on the one hand, social robots are utilitarian systems: they can do household tasks. On the other hand, social robots are hedonic systems: they offer interaction possibilities so as to be able to build long-term relationships with their users. Therefore, it is important to consider hedonic factors as well as utilitarian factors, to get a more complete overview of the important factors in the process of acceptance of social robots.

Several studies with zoomorphic robots were conducted in the last few years. However, it is striking that only a few scholars specifically focused on the acceptance of robots by users [Heerink, Kröse, Evers and Wielinga, 2006; 2008] [De Ruyter et al., 2005]. Research with the iCat showed that a more socially intelligent robot would be more likely to be accepted by users [De Ruyter et al., 2005]. [Heerink et al., 2006] found that there is an influence of perceived social abilities on acceptance of the iCat. [Looije et al., 2006] found that the socially intelligent iCat was preferred by most of the participants. [Heerink et al., 2008] also found that enjoyment influences the intention to use the iCat and this increases the likelihood that people will actually use the iCat. Playfulness is assumed to be an important factor concerning acceptance of robots as well. [Leite et al., 2008] showed

that an iCat with a more playful character helped users to have a better perception of a game played with the iCat.

Interacting with robots seems to be a social activity. When interacting with humanoid robots for the first time, people seem to approach them in groups, e.g. [Shiomi et al., 2006, p.311] "Its name-calling behaviour attracted many visitors. They tried to show the RFID tags embedded in the nameplates to the robot. Often, when one visitor did this, several other visitors began showing their nametags too, as if they were competing to have their names called." [Weiss et al., 2008] studied users' first everyday life experiences with a mechanoid robot. Their results showed that mechanoid robots were also approached in groups. Every time someone tried to interact with the robot via the touch screen, minimally 10 others became curious and started to interact with the robot as well. We are curious to find out whether zoomorphic robots are also approached in groups when first time interactions take place.

Furthermore, robots also seem to be a topic of conversation. People tend to talk about robots with each other. E.g. [Robins, Dautenhahn, Te Boekhorst and Billard, 2004] showed that a robotic doll was used by autistic children as a mediator to interact with adults around them (investigators and carers). [Fujita, 2004] also found that when AIBO was present in a group of children, there were mutual interactions among the children, involving eye contact and some conversations. Results concerning the treatment of older people with dementia also showed that interacting with a zoomorphic robot leads to more communication with others (residents and caregivers) [Shibata, Wada and Tanie, 2003; 2008; 2009] [Wada et al., 2004] [Wada et al., 2005] [Wada et al., 2006] [Wada, Shibata and Kimura, 2008] [Shibata & Wada, 2006; 2007] [Kidd, Taggart and Turkle, 2006]. We are curious to find out whether zoomorphic robots also seem to be a topic of conversation in this study.

[Serenko, Bontlis and Detlor, 2007; Serenko, 2008] showed that personal interest in technology (PIIT) is also an important factor in the acceptance process of zoomorphic robots. [Serenko et al., 2007] [Serenko, 2008] did not find a relationship between PIIT and perceived usefulness. Thus, more innovative people do not necessarily find new technologies more useful. [Serenko, 2008] did find a relationship between PITT and perceived enjoyment, suggesting that the more interest people have in new technologies, the more enjoyment is perceived while using new technologies.

In conclusion, several factors appear to play an important role in the acceptance and usage of zoomorphic robots besides the utilitarian factors of the Technology Acceptance Model. Hedonic factors such as perceived enjoyment, perceived playfulness and personal interest in technology seem to be important factors to consider as well, when trying to understand the acceptance of social robots. Social factors such as approaching robots in groups and communicating about robots with family and friends should also be taken into account.

1.2 Long-term Relationships with Social Robots

Long-term relationships between humans and robots are assumed to be very important in the acceptance process of robots. A lot of studies were conducted studying long-term relationships with zoomorphic robots, such as AIBO, a robot resembling a dog, e.g. [Friedman, Kahn Jr and Hagman 2003] [Kahn Jr, Friedman and Hagman 2002] [Kahn et al., 2004] [Fujita, 2004] [Bartneck et al., 2007] [Turkle et al., 2006] [Tamura et al., 2004] [Stanton et al., 2008], Phyno, a penguin-like robot [Lee et al., 2009], and Paro, a seal robot used for animal assisted therapy with older people suffering from dementia, e.g. [Shibata et al., 2003] [Shibata et al., 2008; 2009] [Wada et al., 2004][Wada et al., 2005][Wada et al., 2006][Wada et al., 2008] [Wada & Shibata, 2006; 2007] [Kidd, et al., 2006].

[Wada et al., 2005] [Wada & Shibata, 2006] and [Kidd et al., 2006] studied the possibility of robot therapy among older people. [Wada et al., 2005, p.2788] describe the example of an older woman, who talked to Paro, after not interacting with him for a month because she was in hospital for treatment: "I was lonely Paro. I wanted to see you again." Participants in the study of [Wada & Shibata, 2006] stated that they felt better after Paro was introduced in their nursing home. They felt as if they had a new playmate and felt less lonely. These results indicate that relationships of humans with a zoomorphic robot such as Paro could be established. This was also stated by [Kidd et al., 2006, p. 3]: "Some residents expressed a special attachment to Paro. They spoke to it like it was a pet, gave it names and engaged it in (one-sided) conversations [….] These users generally began a relationship with Paro in which they saw it as dependent of them. Very often they are/were pet owners."

There seem to be two different categories of how people interact with robots: either they see robots as artificial/as a machine, or they love and nurture them and build a relationship with it, as in the examples of [Wada & Shibata, 2006] and [Kidd et al., 2006]. For example, [Turkle et al., 2006] found that an older man interacted with a robotic doll as if it were his ex-wife. Another older man saw the robotic doll as an interesting artefact and he slapped it just to see what would happen. The man who saw the robotic doll as an artefact talked about the robot when interacting with the researchers, while the man who saw the robotic doll as if it were his ex-wife talked directly to the robot itself. A girl studied by [Turkle et al., 2006, p.351], nurtured an AIBO all the time and saw AIBO as a living being "Oh that is what my dog does when he wants attention… I think it might be sleeping. Or just stretching in a different way than a normal dog would." Another example was found in the study of [Lee, Yamazaki and Helal, 2009], who studied long-term relationships with Phyno. They found that subjects interacted differently with Phyno: they interacted with it as if Phyno was either a machine or a real creature. Thus, to be able to study whether long-term relationships with zoom-

orphic robots occur, the interaction between human and robot should be taken into account.

1.3 Research Questions

The purpose of this study is to get more insight in how people use zoomorphic robots in their homes and whether people are able to build long- term relationships with them. Consequently, the main research questions of this study are:

"How are zoomorphic robots used by people in their domestic setting?"

"Which factors play a role in building and maintaining a long-term relationship with zoomorphic robots?"

Until now very few studies, e.g. [Wada et al., 2004] [Wada et al., 2005] [Wada et al., 2006] [Wada et al., 2008] [Kidd, et al., 2006] studied the usage of zoomorphic robots over a longer period of time. We think that a longer time period is necessary to study whether people can build long-term relationships with social robots. The participants will interact with the rabbit for ten days in three different studies. The number of participants will increase cumulatively over time, aiming to retain the participants from the previous study. Therefore there are three during the first study, six during the second study, and nine during the third study. This study is also novel in its combination of factors that are assumed to be important for the acceptance and building of relationships: besides the often studied utilitarian factors of TAM, hedonic factors, social factors and personal interest in technology will also be taken into account.

2 Method

The zoomorphic robot used in this study is Violet´s Nabaztag, type Nabaztag:tag: a rabbit-shaped Wi-Fi enabled ambient electronic device that can connect to the internet to process specific services via a server located at www.nabaztag.com. The Nabaztag is able to receive predefined spoken commands, but it is not able to understand natural language. It has no mechanisms of learning or memory. Through its programmability, however, the Nabaztag can serve as a robotic user interface to intelligent applications that make use of external sensors and programs. More information about the Nabaztag can be found on the website www.nabaztag.com.

2.1 Procedure

The Nabaztag was installed for 10 days in the participants' homes. It was installed in a place where participants passed it when leaving the house. The set-up consisted of the Nabaztag, a microphone and a video camera which was installed above the Nabaztag, as can be seen in figure 1.

The participants were told that the goal of this study was to help the participants lead a healthy lifestyle. Therefore the function of the system was to ask the participant if they were sticking to an activity plan which they had devised themselves, to ask them to reflect on how they were feeling after a day which had involved some activity, and to ask them to weigh themselves to keep track of their own weight as an indication of their long term health and fitness. The participants interacted with the Nabaztag using buttons for 'yes' and 'no'. The Nabaztag could also provide participants with a weather report and could pass on messages from the researchers. The Nabaztag initiated different conversations at five different times of the day, namely (1) at the first appearance of participants in the morning, (2) when participants were going out of the house, (3) when participants were coming home (4) at a designated time after the last planned activity of the day and (5) when participants were receiving a message from the researchers. At the beginning of every interaction, the Nabaztag asked participants to press a button to give permission to be videoed at that point in time. These videos will be analyzed at a later time.

Figure 1 Set-up of the Nabaztag

2.2 Participants

Three respondents participated in this first study. The aim is that they will also participate in the second and third study. All participants were citizens of the United Kingdom, female and older than 50 years of age. The educational level of the participants differed: one had a bachelor's degree, one had a master's degree and one was in formal education until the age of 16. Two of the participants were employed, one was retired. Two of the participants lived alone, and one lived with her husband. Participants were asked about their interest in technology and all participants were fairly interested. They all belonged to the early majority in the adoption process of technologies [Rogers, 1995]. After the interviews were completed, participants received £20 as a compensation for energy costs incurred during the study.

2.3 Interview Scheme

After the 10 day interaction period, the participants were interviewed about their experiences with the Nabaztag. All interviews were audio-recorded with permission of the participants. The interview was semi-structured and

the primary questions were the topics addressed in Table 1. After answering the primary question, secondary questions were asked to clarify the answers or to ask questions about topics the participant did not mention.

Table 1 Used topics/categories during the interviews

Topics	Categories
General use of Nabaztag	Intention to usage (Lee, Lee and Lee, 2006) Usefulness (Lee et al. 2006) Usage (Lee et al. 2006 Expectations Health exercises Evaluation of the possibilities of the Nabaztag (usefulness of design)
Communication with the Nabaztag (information, appearance, interaction)	Perceived enjoyment (Davis, Bagozzi and Warshaw,1992; Serenko, Bontis and Detlor, 2007; Serenko, 2008) Perceived playfulness (Kim & Moon, 2001; Ahn, Ryu and Han, 2007)
Relationship development with the Nabaztag	Trust (Rau, Li and Li, 2009) Likeability (Rau et al. 2009) Source credibility (Rau et al. 2009) Appearance (and the uncanny valley) Relationship building Novelty effect
Social factors (family/friends)	Subjective norm (Lee et al. 2006) Self-identity (Lee et al. 2006)
Personal interest in technology Demographic variables	Personal interest in technology (Serenko et al. 2007; Serenko, 2008)

2.4 Data Analysis

After the interviews, the recordings of the interviews were transcribed verbatim. After transcription, simple serial indexing was used to analyse the data. Data was categorized via the used categories and the literal transcribed answers of the participants were added to these categories [Mason, 2002].

3 Results

In this article we only present the findings of our first round of analysis, namely the analysis of the simple serial indexing of the data. In a later, more extended, article the results of a cross-sectional analysis and a video analysis will also be described, including the influence of PIIT.

3.1 Usage in Home Environments

Regarding the utilitarian factors, the participants did not find the Nabaztag a very useful device in general. One reason was that the conversations of the Nabaztag were limited to the activity plan ("it said the same things all the time"). However, participants found the Nabaztag easy to use, except for the usage of the conversation buttons. Still, all participants would like to continue using the Nabaztag in study 2 and 3.

Looking at the hedonic factors, two participants did not find it enjoyable to use the Nabaztag due to technical problems and the limited conversation abilities of the Nabaztag (e.g., that it repeated messages). The third participant thought that it was fun to use the rabbit. None of the participants perceived playfulness when using the rabbit.

Regarding the social factors, participants did discuss the Nabaztag and tended to show (pictures of) the zoomorphic robot to family and friends. "I talked with a few people about it. Not many." "[….] I did show one or two a photograph so that they'd know what I was talking about."

3.2 Long-term Relationships

Two of the three participants did not build a relationship with the Nabaztag. They also did not give the Nabaztag a name. One participant built a relationship with the Nabaztag, giving it the name Harvey, and finding the rabbit enjoyable to use. She described the relationship between herself and the Nabaztag as: "He asked the questions, I answered them." Although she did build a relationship with it, she did not see the Nabaztag as a friend. "No, I just got used to this, he was a presence. He's a man-made presence or even a women-made presence, in my kitchen, who was doing a job of research. I always knew that that's what it was."

4 Discussion

The results showed that regarding the utilitarian factors, the participants did not find the Nabaztag useful. However, they found the Nabaztag easy to use, except for the buttons that were used to communicate with the Nabaztag. No evidence was found at this stage to indicate that hedonic factors, enjoyment and playfulness, were of importance for the acceptance of robots. Even though the Nabaztag was not perceived as useful and enjoyable, all participants indicated that they would like to continue to use the Nabaztag in the second and third study.

There was one indication that hedonic factors are of importance in building a relationship with the Nabaztag. The results showed a relationship between hedonic factors (in this case enjoyment) and building a relationship with the Nabaztag, since the only participant who perceived hedonic factors was able to build some kind of relationship with the social robot. More evidence is needed to confirm whether hedonic factors are of importance for the acceptance of the Nabaztag, as shown in the results of the studies of [Heerink et al. 2006; 2008] [De Ruyter et al., 2005] [Leite et al., 2008].Therefore, these issues will be further explored in the second and third study.

Social factors did seem to be of importance for the acceptance but not for relationship building with the Nabaztag. The results showed that the Nabaztag was discussed with family and friends. This finding is consistent with the results of [Robins et al., 2004] [Shibata et al., 2003; 2008; 2009] [Wada et al., 2004] [Wada et al., 2005] [Wada et al., 2006] [Wada et al., 2008] [Wada & Shibata, 2006; 2007] and [Kidd et al., 2006]. The results also showed that participants tended to show pictures of the Nabaztag to family and friends. Showing pictures to family and friends is a finding that was not reported in

other studies. This might suggest that the participants found the Nabaztag important enough to show it to their family and friends. This could imply that they saw it as more than an article of use. Discussing or showing the Nabaztag to family and friends does not seem to be related to relationship building in this study, since all participants talked about it and showed the robot to others. This could indicate that in a later stage, after some technical improvements, all participants could build a relationship with the Nabaztag.

Another interesting finding was that there seems to be a relationship between naming the rabbit and building relationships. The results show that only one participant was able to build a relationship with the Nabaztag. This participant was also the only one who gave the Nabaztag a name. This indicates that giving the Nabaztag a name could be related to relationship building. The amount of participants who were able to build a relationship with the robot is consistent with earlier results, namely approximately one third of the participants were able to build a relationship with a robot [Kidd et al., 2006].

A limitation of this study was that the goal presented to the participants, to help the participants lead a healthy lifestyle, was not accomplished due to technological problems and the simplicity of the system. Improvement should be made for the next iterations. Another limitation was the small number of participants. But small, qualitative studies are an essential step to larger studies. Another limitation was that we could not study whether people approached the Nabaztag in groups when interacting with it for the first time due to the fact that two of the three participants lived alone.

5 Conclusion

This study yielded interesting insights which will be further explored in our next two studies. Our focus of attention in these studies will be (1) to establish whether hedonic factors and social factors are important in accepting zoomorphic robots, (2) to explore the relationship between hedonic social factors and relationship building with zoomorphic robots and (3) to explore the relationship between name-calling and relationship building with zoomorphic robots.

6 Acknowledgements

This study was supported by funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement no.231868 and project name, Social Engagement with Robots and Agents (SERA).

References

[Ahn, T., Ryu, S. and Han, I., 2007] The impact of Web quality and playfulness on user acceptance of online retailing, Information & Management: 263-275.

[Bartneck, C., Suzuki, T., Kanda, T., Nomura, T., 2007] The influence of people's culture and prior experiences with Aibo and their attitudes towards robots. AI & Soc 21: 217-230.

[Davis, F. D., Bagozzi, R.P. and Warshaw, P.R., 1992] Extrinsic and intrinsic motivation to use computers in the workplace. Journal of applied social psychology 22(14): 1111-1132.

[Dautenhahn, K., 2002] Design spaces and niche spaces of believable social robots. 2002 IEEE Int. Workshop on Robot and Human Interactive Communication. Berlin, Germany, IEEE Press: 192-197.

[Friedman, B., Kahn Jr., P.H. and Hagman, J., 2003] Hardware companions? - What online AIBO discussion forums reveal about the human-robotic relationship. CHI letters 5(1): 273-280.

[Fujita, M., 2004] On activating human communications with pet-type robot AIBO, Proceedings of the IEEE 92(11).

[Heerink, M., Kröse, B., Evers, V. and Wielinga, B., 2006] The Influence of a Robot's social abilities on acceptance by elderly users. Amsterdam, Hogeschool van Amsterdam.

[Heerink, M., Kröse, B., Evers, V. and Wielinga, B., 2008] Enjoyment, intention to use and actual use of a conversational robot by elderly people. HRI'08. Amsterdam, the Netherlands: 113-119.

[Heijden, v. d., H., 2004] User acceptance of hedonic information systems. MIS Quarterly 28(4): 695-704.

[Kahn Jr., P. H., Friedman, B. and Hagman, J., 2002] 'I care about him as a pal": conceptions of robotic pets in online AIBO discussion forums, CHI2002: 632-633.

[Kahn Jr., P. H., Freier, N.G., Friedman, B., Severson, R.L. and Feldman, E.N., 2004] Social and moral relationships with robotic others. The 2004 IEEE International Workshop on Robot and Human Interactive Communication, Kurashiki, Okayama Japan: 545-550.

[Kidd, C. D., Taggart, W. and Turkle, S., 2006] A sociable robot to encourage social interaction among the elderly. International conference on robotics and automation 2006: 3972-3976.

[Kim, J.-W. and Moon, Y.-G., 2001] Extending the TAM for a World-Wide-Web context. Information & Management 38: 217-230.

[Lee, Y., Lee, J. and Lee, Z., 2006] Social influence on technology behaviour: self-identity theory perspective. The DATA BASE for Advances in Information Systems 37(2 & 3): 60-75.

[Lee, D., Yamazaki, T. and Helal, S., 2009] Robotic companions for smart space interaction, Pervasive Computing: 78-84.

[Leite, I., Martinho, C., Pereira, A. and Paiva, A., 2008] iCat: an affective game buddy based on anticipatory mechanisms (short paper). Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems, Estoril, Portugal: 1229-1232.

[Looije, R., Chossen, F. and Neerinckx, M.A., 2006] Incorporating guidelines for health assistance into

a socially intelligent robot. The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06) Hatfield, UK: 515-520.

[Mason, J., 2002] Qualitative Researching second edition, Sage: London, Thousand Oaks, New Delhi.

[Rau, P.L., Li, Y. and Li, D., 2009] Effect of communication style and culture on ability to accept recommendations from robots, Computers in Human Behaviour: 587-595.

[Robins, B., Dautenhahn, K., Boekhorst, te, R. and Billard, A., 2004] Robots as assistive technology – Does appearance matter? Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication, Kurashiki, Okayama Japan.

[Rogers, E.M., 1995] New product adoption and diffusion, Journal of costumer research 2 (march), 290-301

[Ruyter, de B., Saini, P., Markopoulos, P. and Breemen, van A., 2005] Assessing the effects of building social intelligence in a robotic interface for the home, Interacting with computers 17: 522-541.

[Social Engagement with Robots and Agents (SERA) EU project FP7-231868]

[Serenko, A., Bontis, N. and Detlor, B., 2007] End-user adoption of animated interface agents in everyday work applications. Behaviour & Information Technology 26(2): 119-132.

[Serenko, A., 2008] A model of user adoption of inter face agents for email notification. Interacting with computer: 461-472.

[Shibata, T., Wada, K. and Tanie, K., 2003] Statistical Analysis and Comparison of Questionnaire Results of Subjective Evaluations of Seal Robot in Japan and U.K. The 2003 IEEE International Conference on Robotics and Automation, Taipei, Taiwan: 3152- 3157.

[Shibata, T., Wada, K., Ikeda, Y, and Sabanovic, S., 2008] Tabulation and analysis of questionnaire results of subjective evaluation of seal robot in 7 countries. The 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Germany, 689-694.

[Shibata, T., Wada, K., Ikeda, Y. and Sabanovic, S., 2009] Cross-cultural studies on subjective evaluation of a seal robot. Advanced robotics 23, 443-458.

[Shiomi, M., Kanda, T., Ishiguro, H. and Hagita, N., 2006] Interactive humanoid robots for a science museum. HRI'06. Salt Lake City, Utah, USA, 305-312.

[Stanton, C. M., Kahn Jr., P.H., Severson, R.L., Ruckert, J.H. and Gill, B.T., 2008] Robotic animals might aid in the social development of children with autism, HRI'08, 271-278.

[Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, A., Higashi, Y., Fujimooto, T. and Nakajima, K., 2004] Is an entertainment robot useful in the care of elderly people with severe dementia? Journal of gerontology: MEDICAL SCIENCES 59a(1), 83-85.

[Turkle, S., Taggart, W., Kidd, C.D. and Dasté, O., 2006] Relational artefacts with children and elders: the complexities of cyber companionship, Communication Science 18(4): 347-361.

[Wada, K., Shibata, T., Saito, T. and Tanie, K., 2004] Effects of robot-assisted activity for elderly people and nurses at a day service centre. Proceedings of the IEEE 2004, 1780-1788.

[Wada, K., Shibata, T., Saito, T., Sakamoto, K. and Tanie, K., 2005] Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. IEEE International conference on robotics and automation, Barcelona, Spain, 2785- 2790.

[Wada, K. and Shibata, T., 2006] Robot therapy in a care house - results of case studies. The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06). Hatfield, UK (September 6-8, 2006), 581-586.

[Wada, K., Shibata, T., Musha, T. and Kimura, S., 2006] Effects of robot therapy for demented patients evaluated by EEG, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005. (IROS 2005), 1552- 1557.

[Wada, K. and Shibata T., 2007] Living with seal robots - its socio-psychological and physiological influences on the Elderly at a care house. IEEE Transactions on robotics 23(5), 972-980.

[Wada, K., Shibata, T., Musha, T. and Kimura, S., 2008] Robot therapy for elders affected by dementia. IEEE engineering in medicine and biology magazine July/August, 52-60.

[Weiss, A., Bernhaupt, R., Tscheligi, M., Wolherr, D., Kuhnlenz, K. and Buss, M., 2008] A methodological variation for acceptance evaluation of human-robot interaction in public places. The 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany, 713-718.

Conversation in Context: what should a robot companion say?

Peter Wallis, Viktoria Maier
Dept of Computer Science,
The University of Sheffield,
Sheffield, S1 4DP, UK
firstInitial.surname@dcs.shef.ac.uk

Sarah Creer and Stuart Cunningham
ScHARR (Health Services Research),
The University of Sheffield,
Sheffield, S1 4DP, UK
firstInitial.surname@sheffield.ac.uk

Abstract

Language as used by humans is a truly amazing thing with multiple roles in our lives. Academics have tended to focus on the way languages convey meaning, and disciplines that come new to the problem such as computer science tend to start with reference semantics and progress to models of meaning that look mathematical and hence solidly academic. Language as used is however beautifully messy. People sing, they lie and swear, they use metaphor and poetry, play word games and talk to themselves. Is there a better way to look at language? Interdisciplinary research is hard not only because each discipline has its own terminology, but also because they usually have different interests. Those of us interested in spoken language interfaces (computer science) however have a shared interest with applied linguistics in how language works in situ. This paper outlines a theory about how language works from applied linguistics and shows how the theory can be used to guide the design of a robot companion.

1 Introduction

In 2005 and 2006 some of us were involved in a workshop series on computers and abuse. Our motivating interest was in why people swear at chat-bots. This is not some minor fluctuation - de Angeli looked at transcripts from Jabberwacky and found 20% of the *words* were abusive [de Angeli, 2005]. Indeed it seems this abuse is not species specific. In experiments with an Aibo and dogs, there is dramatic footage of the dog throwing the machine across the room. It seems that we animals do not like machines. What is happening? In one of the workshops it was proposed that the abuse we observe might be part of some unconscious mechanism that enables intra species cooperation [Wallis, 2005]. In the same way as termites unthinkingly (presumably) follow rules that result in large complex artifacts, might there be simple rules of human behaviour that in some way enable our more grand achievements. It turned out that a description of such rules can be found in applied linguistics.

Conversation Analysis (CA) is a methodology with a strong commitment to naturally occurring data and the ethnomethodological variants have strong links with anthropology. Its techniques are aimed at noticing and explaining the everyday - the things that we do without thinking. CA is usually associated with the very low level details of conversation - the nature of turn taking, the structure of openings, the notion of adjacency pairs and so on - but Seedhouse [Seedhouse, 2004] sums up the finding of CA over the years with the macro observation that an utterance in a conversation either goes "seen but unnoticed, noticed and accounted for, or risks sanction." In the case of chatbots, this sanction takes the form of swearing.

The theory is that language works in the first instance in much the same way as we computer scientists think it does with words referencing things and action in the world. In language in use, your conversational partner (CP) is likely to produce an utterance that goes **seen but unnoticed** providing answers to questions, returning greetings and giving explanations as you expect.

When that fails - when your CP's response is in some way out of the ordinary - you, the listener, notice the utterance and work very hard to find an explanation for what he or she said. An utterance becomes **noticed and accounted for**. It is worth quoting at length the Eggins and Slade [Eggins & Slade, 1997] description of just how much emphasis we humans place on sequential relevance:

> Consider, for example, the two turns at talk below:
>
> A What's that floating in the wine?
> B There aren't any other solutions.
>
> You will try very hard to find a way of interpreting B's turn as somehow an answer to A's question, even though there is no obvious link between them, apart from their appearance in sequence. Perhaps you will have decided that B took a common solution to a resistant wine cork and poked it through into the bottle, and it was floating in the wine. Whatever explanation you came up with, it is unlikely that you looked at the example

and simply said "it doesn't make sense", so strong is the implication that adjacent turns relate to each other.

This is where, in computer science terms, things get untidy and where, for instance, the need for context is found. The CA people, interestingly, note that context is somehow *selected* by what is said.

Finally, if you fail to account for your CP's utterance, then your CP **risks sanction**. This may take the form of verbal abuse and swearing, but it can also mean a telling off if there is a power relation or simply walking away and avoiding them in future. For chat-bots, its swearing.

2 Providing an accounting for

This notion that each conversational partner's utterance is either seen but unnoticed, noticed and accounted for, or risks sanction, provides an explanation for the considerable interest in the literature in having conversational characters that are *consistent*. It has been claimed that clarity and consistency are among the most important features in an agent [Norman, 1988; Tognazzini, 1992; Winograd & Flores, 1987]. It allows people to predict what will happen when they engage with the system lightening cognitive load [Fiske & Taylor, 1991], and point out that consistency also means consistency with a user's expectations [Tognazzini, 1992, 250,251]. Thus, good agent design should work with the already existing set of expectations that a person brings to his/her interaction with the object, creating intuitive "affordances" [Norman, 1988]. Nass & Brave [Nass & Brave, 2005] point out that inconsistency leads to difficulties in perception and processing. These claims are intuitively sensible, but the realisation that people work hard to account for an agents action provides the mechanism by which consistency helps. A consistent character is one that provides an accounting for its behaviour.

This notion of seen but unnoticed, noticed and accounted for, or risks sanction, also provides a useful framework for analysing system performance. As a classic example of a machine's interaction eliciting sanctions, here is a conversation from the 1996 Loebner Competition:

PROGRAM	G'day mate.
JUDGE	G'day. How's it going?
PROGRAM	My mother once said "How it is going is a mystery." She was a bit nuts.
JUDGE	Uh-huh
PROGRAM	Are you still learning to speak, my friend?
JUDGE	NO, no. But your anecdote took me a bit by surprise.
JUDGE	How was your morning?
PROGRAM	Oh well, that's a pity!
JUDGE	I'm sorry; what's a pity?
PROGRAM	Apologies accepted.
JUDGE	I feel compelled to insult you and we've only been talking a minute.
PROGRAM	We are bloody well talking, aren't we?
JUDGE	Oh baby. We sure are. So here are some keywords for you: shit mother baseball.
	...

This was the winning entry from that year and of course there are more recent and better funded projects. The DARPA Communicator programme was a brilliant success in that it reflected just how much we do not know about language in use. The participating organisations had access to world experts on spoken language interfaces, and effectively unlimited resources, but failed to get anywhere near the vision as set out by the programme coordinators. The observation is that these systems did mostly well when the human stuck to answering the system's questions and the automatic speech recognition (ASR) worked. That is, we know how to handle the seen but unnoticed. When the grounded knowledge was not as the script designers expected however, the system did not have the world knowledge to understand the user's move to repair the situation and the system would produce a response that the user could not account for. The result was sanction. In Communicator, swearing at the system is rare (surprisingly) for reasons that can only be guessed at. One can however see (hear) sanction happening in the recordings with either the user expressing exasperation or simply hanging up the phone, but it is also reflected in the post call survey in which a significant proportion of users did not want to "use the system on a regular basis" [Wallis, 2008].

As someone with connections with the dialog systems community it is tempting to blame the speech recognition technology but there is a growing body of evidence that word error rates are not the problem [Wallis *et al.* , 2001; Skantze, 2007]. The problem is not "trouble in text" such as failed ASR or a lack of world knowledge, but rather people get annoyed with spoken language interfaces because the systems we produce fail to repair the trouble. Trouble happens all the time in human-human communication but this is not annoying in itself as long as one's conversational partner is observably trying to help. That is, as long as your CP's utterances can be accounted for, you don't even notice that trouble occurred.

The mechanism for accounting for can be both tactical and strategic. Eliza and Parry were very successful in that user satisfaction was very high compared to modern day systems. The mechanism was strategic in those systems in that they provide an accounting for their behaviour – in the first case because the role of psychologist accounts for the endless stream of personal questions, and in the second because being paranoid accounts for the system's odd responses and interests. At the tactical level, at least one of the systems in the Communicator programme would, if the conversation got recognisably confused, say the "network connection seems to be down" and ask the user to try again later, effectively accounting for its fail-

ure to provide a flight booking. Emotion too can play a tactical role. Eugene was a virtual cuttlefish that expressed emotion through colour [Wallis & Wilks, 2005]. At the strategic level, Eugene's "personality" was rather vain and he liked to talk about his pretty colours. Stuck for something to say Eugene would ask you what you thought of his colours. This otherwise rather strange conversational gambit worked because his persona accounted for the behaviour and so it did not seem strange at all. If you were rude to Eugene, he would change the subject giving the impression that he was offended and hence accounting for the change of subject. Once again, the accounting for is so effective that for us humans (as opposed to us researchers) it is hard to notice what the issue is.

It seems a natural language interface can get away with many things, including non understanding and poor ASR, if it is set up so that the user can *account for* its behaviour. In our tool box of techniques are persona, role and emotion, but in human-human dialog we also find explicit discussions to account for possibly unexpected behaviour. In a set of experiments based around booking cars out of the Division's car pool [Wallis *et al.* , 2001], there were explicit discussions to account for delay (often waiting for the computer to do something) and indeed about high level plans. The subject would describe to the caller how she was going to handle their call, followed by "is that ok?" - which was not really a question but was an opportunity to provide more information that would change her plan. One might have said "No, just do ... ", but such an unaccounted for move would certainly risk sanction.

Finally, note that although a person will work very hard to find an interpretation of their conversational partner's utterance, this requires that the person is committed to the conversation. I will work hard to figure out why a CP said what he or she said, in the way it was said, but to do that requires that I know that I am being addressed. This is not an issue for a chat-bot on a website or a system set up for experiments in a laboratory, but becomes a significant issue for an interactive artifact that is permanently in someone's home.

3 SERA

The SERA project is an EU programme funded through the FP7 ICT call 3 on cognition and robotics to look at "social engagement with robots and agents." The focus has turned out to be on robots rather than embodied conversational agents and the robot of choice is a Nabaztag. The Nabaztag is a commercially produced talking head from Violet in the style of kismet, the iCat, or indeed Foust's talking skull. It is a stylised rabbit with expressive ears and a set of multi colour LEDs (see Figure 1) and is marketed as the world's first internet enabled talking rabbit. The rabbit connects to the Violet server via a wireless router and can run several applications including receiving SMS messages, weather reports, tai chi, and streaming selected radio or blog sites.

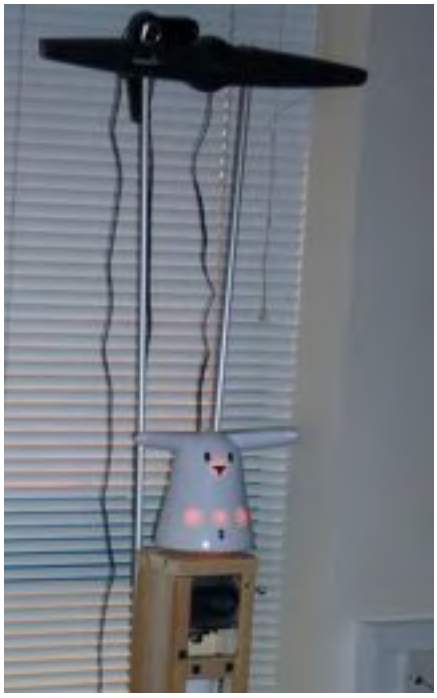

Figure 1: The Nabaztag on a pedestal (with array microphone, web-cam, and a PIR sensor)

Our aim is to study long term relationships between people and robot "companions" and the intention was to put a Nabaztag in an older person's home and see what happens. This is not as straight-forward as it may first appear as older people are not the intended market niche for this product. Mival et al. [Mival *et al.* , 2004] find that products such as the Furby and Aibo tend to end up "in the back of the cupboard with the batteries out," and our aim was to study the way relationships with a machine develop over time. They suggest that a synthetic companion needs to have a perceived use even if that use is not really beneficial. Unless one likes technology for its own sake, the user needs a reason for having the thing on.

The use of the robotic agent we settled on in this case was a health and exercise application. Following the British Heart Foundation literature [Lewin *et al.* , 2005] for individuals recovering from heart failure, the participants were asked to write an **activity plan**, and the intention was to have the robot rabbit help them stick to it. The challenge for people recovering from a heart attack is to do some exercise, but not too much and this can be quite complex when housework or indeed going up stairs is significant. This actual application is too fraught with responsibilities and ethical issues and so our actual subjects were recruited from a pool of healthy older people through community groups such as Aged Concern and Help the Aged. Dialogs for the rabbit were scripted based on the transtheoretical model (TTM) of behaviour change [Prochaska & Velicer, 1997] which is discussed in more detail below. The rabbit was also designed to provide short small talk type interactions, provide a weather report and pass on messages from the re-

searchers. The SERA programme of work takes the form of three iterations of field studies. At each stage the system is deployed in users' houses and remain with them for ten days, with the observations from each iteration informing developments for the following iteration.

Iteration one was intended primarily as a trial of the set-up, our subject recruitment and data collection policies. The first finding was that the technology has its limitations. Interestingly we failed to get the speech recognition software working in a way that could be incorporated in a dialog system. Indeed it has emerged recently that when NASA use Dragon Naturally Speaking for dialog systems [Hieronymus, 2009], there are "secret switches" that are used to reduce response time to less than half a second. Without access to these modifications, we modified our dialogs to use yes/no buttons mounted on the stand for the rabbit in iteration one, and will use a different speech recognition system for the remaining deployments.

Our aim was for 6 interactions per day for each of our 3 subjects in iteration 1. One system failed completely, and the other two systems came no where near that number of recorded interactions. Inspite of this, there are a surprisingly large number of "interesting" recordings.

3.1 Analysis

Following on from the previous work on abuse, our analysis focused on situations in which things go wrong, and our methodology was based on techniques from Conversation Analysis, or CA. In particular we followed the principle that "no detail is too small for analysis, but not every detail necessarily forms part of the analysis," and we embraced the ethnomethods principle that people give off signals that (as humans, not researchers) we are usually very good at reading. The assumption is that as humans we can recognise when there is "trouble in text" in the video, and our job as researchers is to come up with explanations of how that trouble came about. There is then the engineering issue of what can be done about it.

Evidence from the data collected in the first iteration shows that, specifically, the activity-related interactions did not promote a positive relationship between the user and the device. Examples include eye rolling and dismissive hand gestures. The observable physical action does not *mean* anything, but the significance of such actions can be *read off* in context from the video by any untrained human. At some point in the future we hope to annotate the video with some transcription scheme that captures this level of description.

An explanation for this lack of perceived usefulness can be made in terms of the transtheoretical model mentioned above. The TTM consists of five stages of change: precontemplation, contemplation, preparation, action and maintenance. The activity plan task as designed for the first iteration was intended to persuade, and clearly fits in with the preparation and action stages of the model. However, by responding to the recruitment advert, the individual has moved beyond the stages of precontemplation and contemplation and, evidence suggests, our subjects are in the maintenance phase of this model. That is, they have a set activity plan which they are currently following. In the post interviews this was clearly stated:

P1: "But I never forget that I am going to Aqua [aerobics], because there is only so many sessions a week."

P2: [the rabbit] "asked me if I have had done my activity, which I do if I want to and I very rarely don't. And if I don't do there is a very good reason. I didn't need to be told."

What is more, from the system logging data, when the rabbit asked "Did you stick to the amount of activity that was in your activity plan?", there were zero occurrences of a 'no' response. The users are already leading what they consider to be a healthy active lifestyle, and do not see themselves as within the group of individuals at whom this device is aimed.

So if they did not see the system as useful to them, what was their motivation? The answer, obviously, is that they are participating in the project to help others via this research. Again from the interview data:

P1: "the whole point of this is to see if, you know, us old 'uns just sit on the settee all day long doing nothing isn't it? 'Cause some of them do."

Although the subjects were happy to have the system in their home for ulturistic reasons, the task it performs is not consistent with the expectations or requirements of the user and therefore not useful or relevant to them. It is on its way to the back of the cupboard with the batteries out.

4 Re-designing the Interaction

With people not liking the content of the interactions, and with such a low number of interactions recorded, perhaps we should consider a different role for the rabbit. Perhaps a more entertaining role and/or a more dynamic personality would result in more interactions — and more positive ones at that. In the end we decided to stick with the exercise task for two reasons. First, we think we know how to fix the discussions of the activity plan and second, by sticking with the exercise task, we believe we have a way to account for the system's limited knowledge about the world.

4.1 Discussing the Activity Plan

The aim is to tailor the dialogs toward people in the maintenance phase of the TTM of behaviour change. There are many occasions in the video data and interviews where the individuals were keen to talk about how much exercise they did and tell the rabbit more about their activity. The concept of 'regression', the move backwards to a previous stage, is relevant at all stages of the model and the prevention of this will be the focus for the new task. A vital construct in the prevention of regression is that of 'self-efficacy' [Bandura, 1977], the context-specific confidence that an

individual has in their ability to fulfil their goals. Maintaining a high level of self-efficacy will help to keep the individual in the current stage. It is therefore hypothesised that the activity related interactions should effectively provide confidence boosting or confidence maintaining behaviour. Providing a way for this information to be input to the rabbit and feeding that information back in a summary to the participant will remind them of how much they have achieved. This will more accurately match the users' image of themselves and their current stage in the model. It is hypothesised that this will make the information relevant to them and increase the perceived usefulness and therefore engagement with the robotic agent.

4.2 Persona

Ideally our robot companion would have a fully fleshed out human character with a history, likes and dislikes, friends and relations, and be able to talk on any subject from quantum physics to Coronation Street[1]. The critical thing is that whatever conversations the agent has, it must either know the things the human thinks it should know, or provide an *account for* the system's lack of knowledge. The dialogs need "coverage" of the domain and an agent that is happy to talk about anything would require the encoding of vast quantities of common-sense knowledge. This is not only time consuming, but projects such as the CYC project indicate that it is not even possible.

A popular mechanism for limiting the scope is to have the system provide a service and take on a role as a service provider. For instance, If I ring for a taxi, I do not — and cannot — expect the person on the other end of the phone to talk about characters in Coronation Street. For the robot companion, this notion of a constrained role is harder as the system is in the user's space permanently. For iteration 2 the intention is to use persona to account for the limited scope of the dialogs.

In iteration 1 the system was not designed with a particular persona in mind and indeed, if conceived of at all, its role was to pass on information rather than to have personal opinions. For iteration 2 the aim was to play to the subjects' perceptions from iteration 1 and the decision was to make a rabbit that is shy and self-deprecating. This persona fits with the perception of a small animal and with its role as passing on information when asked. It also contributes to accounting for its behaviour in that it sits in the user's house and does not always talk to the user when they are nearby. A more proactive persona may intuitively lead to more interactions and more engagement, but that would require more topics to be covered and is likely to result in the user finding the limits of that knowledge which will require an accounting for.

By making our rabbit shy, the subject will be able to account for the rabbit not saying much and, indeed, being hesitant about joining a conversation. There are however limits and a key notion is that the rabbit should not appear rude. To do that, it needs to respond when addressed. The aim is to enable the rabbit to recognize when it is addressed and provide a minimal response.

Back-channel

When engaged in a conversation the person *not* talking participates by nodding his or her head and making appreciative noises. In Japan the expression "hai" is often treated as meaning yes, but in linguistic circles is often translated as "yes, I have heard you". These *back-channel* expressions are often seen as irrelevant and edited out of transcripts but actually perform a function [Schegloff, 1982]. Our model of how language works suggests that the purpose of such signaling is to indicate that there is a commitment to engage. That is, in order for a listener to bother *accounting for* some apparently irrelevant action on the part of the speaker, he or she needs to be engaged as mentioned in Section 2. This back-channeling signals engagement. On hearing "hai", I know that I am being attended to. Such back channelling is semantics free and as such it is feasible that we can re-produce it with our system.

It is evident from the data that we need to do more signaling of, and pay more attention to, attention. From footage of interactions in iteration 1 it was observed that long pauses between button press and activation of the agent resulted in insecurity and impatience in the user. In human-human communication, such pauses are filled with visual and verbal signalling. Since the facial features are not available, new signals may want to be introduced using the ears and the coloured lights of the nabaztag. Such an approach would introduce its own problems however as new signals may not be easily understood as back-channeling by the user, and it should be noted that if back-channel messages are just slightly off-timing or wrong, the result in an interruption or competition for the conversational floor [Young & Lee, 2004].

"Easter eggs"

As mentioned earlier, it would of course be more interesting to have a social agent which is more versed in the real world, but such an approach is beyond the scope of this project. However, one opportunity made possible by having a shy rabbit is that we can focus on having "islands of knowledge". Other than the primary function of the agent, our rabbit could appear fanatical about something. Such a feature would provide variation, and some positive novelty. As an example consider the possibilities if the rabbit was provided with a (perceived) fanatical interest in Coronation Street. If the user mentions a keyword related to the show such as "Dan" or "kill", the shy rabbit suddenly starts a conversation about the latest happenings on the soap. If our human wanders, and starts talking about a corner shop that is not in the television program, then the rabbit becomes shy again and starts to sound less interested. The specialist domain might not be made explicit and indeed, like in the games industry where the notion of an "easter egg" is common, the rabbit's expert knowledge may be hidden and left to be discovered.

[1] "Coronation Street" is a long running British soap opera.

Giving the robot an obsession would provide a means of discussing novel information. This can be done without committing the rabbit to an understanding of the world at large by making the robot appear shy. Being shy provides an accounting for the rabbits disinterest in discussing anything else.

5 Conclusions

The SERA project addresses the notion of social engagement with robots and agents. To this end we have begun the iterative development of a robot that can help its user pursue a healthier lifestyle through regular exercise. We initially saw the system as a persuader, but after our first iteration think that the role is more of an encourager.

We would like to see more interactions in the next iterations and to do that we will extend the role of the robot into islands of knowledge, the limits of which can be accounted for by the persona of the robot.

6 Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 231868.

References

[Bandura, 1977] Bandura, A. 1977. Self-efficacy: toward a unifying theory of behavioural change. *Psychological review*, **84**, 191–215.

[de Angeli, 2005] de Angeli, Antonella. 2005 (September). Stupid computer! abuse and social identity. *In:* Angeli, Antonella De, Brahnam, Sheryl, & Wallis, Peter (eds), *Abuse: the darker side of human-computer interaction (interact '05).* http://www.agentabuse.org/.

[Eggins & Slade, 1997] Eggins, Suzanne, & Slade, Diana. 1997. *Analysing casual conversation.* Wellington House, 125 Strand, London: Cassell.

[Fiske & Taylor, 1991] Fiske, S. T., & Taylor, S. E. 1991. *Social cognition (2nd ed.).* McGraw-Hill.

[Hieronymus, 2009] Hieronymus, Jim. 2009 (October). *Spoken dialogue systems for space and lunar exploration.* SpandH Seminar.

[Lewin *et al.* , 2005] Lewin, B., Pattenden, J., Ferguson, J., & Roberts, H. 2005. *The heart failure plan.* London: The British Heart Foundation.

[Mival *et al.* , 2004] Mival, Oli, Cringean, S., & Benyon, D. 2004. Personification technologies: Developing artificial companions for older people. *In: Chi fringe.*

[Nass & Brave, 2005] Nass, C., & Brave, S. 2005. *Wired for speech: How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press.

[Norman, 1988] Norman, D. A. 1988. *The design of everyday things.* New York: Doubleday.

[Prochaska & Velicer, 1997] Prochaska, J., & Velicer, W. 1997. The transtheoretical model of behaviour change. *American journal of health promotion*, **12**, 38–48.

[Schegloff, 1982] Schegloff, Emanuel A. 1982. Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. *Pages 380–407 of:* Tannen, Deborah (ed), *Georgetown university round table on languages and linguistics 1981.* Georgetown University Press.

[Seedhouse, 2004] Seedhouse, Paul. 2004. *The interactional architecture of the language classroom: A conversation analysis perspective.* Blackwell.

[Skantze, 2007] Skantze, Gabriel. 2007. *Error handling in spoken dialogue systems - managing uncertainty, grounding and miscommunication.* Ph.D. thesis, Department of Speech, Music and Hearing, KTH.

[Tognazzini, 1992] Tognazzini, Bruce. 1992. *TOG on interface.* Addison-Wesley.

[Wallis, 2005] Wallis, Peter. 2005 (September). Robust normative systems: What happens when a normative system fails? *In:* Angeli, Antonella De, Brahnam, Sheryl, & Wallis, Peter (eds), *Abuse: the darker side of human-computer interaction (interact '05).* http://www.agentabuse.org/.

[Wallis, 2008] Wallis, Peter. 2008. Revisiting the DARPA communicator data using Conversation Analysis. *Interaction studies*, **9**(3).

[Wallis & Wilks, 2005] Wallis, Peter, & Wilks, Yorick. 2005. *Companions, emotions and interaction.* Presented at WP6, Humaine workshops. http://emotion-research.net.

[Wallis *et al.* , 2001] Wallis, Peter, Mitchard, Helen, O'Dea, Damian, & Das, Jyotsna. 2001. Dialogue modelling for a conversational agent. *In:* Stumptner, Markus, Corbett, Dan, & Brooks, Mike (eds), *Ai2001: Advances in artificial intelligence, 14th australian joint conference on artificial intelligence.* Adelaide, Australia: Springer (LNAI 2256).

[Winograd & Flores, 1987] Winograd, Terry, & Flores, C. 1987. *Understanding computers and cognition: A new foundation for design.* Reading, MA: Addison-Wesley.

[Young & Lee, 2004] Young, R. F., & Lee, J. 2004. Identifying units in interaction: Reactive tokens in korean and english conversations. *Journal of sociolinguistics*, **8**(3), 380–407.