Abstract—Human-Robot Interaction (HRI) research is here presented into social robots that have to be able to interact with inexperienced users. In the design of these robots many research findings of human-human interaction and human-computer interaction are adopted but the direct applicability of these theories is limited because a robot is different from both humans and computers. Therefore, new methods have to be developed in HRI in order to build robots that are suitable for inexperienced users. In this paper we present a video study we conducted employing our robot BIRON (Bielefeld RObot companionN) which is designed for use in domestic environments. Subjects watched the system during the interaction with a human and rated two different robot behaviours (extrovert and introvert). The behaviours differed regarding verbal output and person following of the robot. Aiming to improve human-robot interaction, participants’ ratings of the behaviours were evaluated and compared.

I. INTRODUCTION

Wehenever technical devices are built for novice users, several design decisions have to be taken with caution. This is also true for social robots which more and more inexperienced users get in touch with. Robots of this kind include toys (e.g. Aibo [1], Furby [2], Lego Mindstorms [3]), robots for cleaning (e.g. Roomba [4]) and healthcare (e.g. Paro [5]). All these robots should be able to fulfill their tasks effectively and users should also like to interact with the systems and want to use them for a long time. To achieve these aims, the robots should be evaluated with the help of potential users. User studies should be part of the development cycle giving useful hints for improvement both of the technical and interaction capabilities of the robot.

Manuscript received February 1, 2008. The work described in this paper was conducted within the EU Integrated Project COGNIRON (The Cognitive Robot Companion’- www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

M. Lohse, M. Hanheide, B. Wrede, and G. Sagerer are with the Applied Computer Science Group, Bielefeld University, 33619 Bielefeld, Germany (phone: 0049-521-1062953; fax: 0049-521-2992; e-mail: mlohse; mhanheide; bwrede; sagerer@techfak.uni-bielefeld.de).

M. L. Walters, K. L. Koay, D. S. Syrdal, and K. Dautenhahn are with University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK (e-mail: m.l.walters; k.l.koay; d.s.syrdal; k.dautenhahn@herts.ac.uk).

A. Green, H. Hüttenrauch, and K. Severinson-Eklundh are with KTH Royal Institute of Technology School for Computer Science and Communication (CSC), SE-100 44, Stockholm, Sweden (e-mail: green; hehu@csc.kth.se; kse@kth.se)

An evaluation approach proposed by [6] is video studies, which provide a supporting method to live user trials. In these studies, interactive robot behaviours are recorded on videotape. In experiments, the video is then shown to many viewers who are asked to rate the behaviours they watch. The method enables researchers to conduct studies with a large sample of participants in a relatively short time. We chose it as one of several to evaluate the domestic robot BIRON (e.g. [7], [8], [9]). This paper presents the results of a study with 200 participants from Germany and Great Britain.

II. METHOD

The practical aim of HRI as a discipline is to design robots that are as effective, efficient and usable as possible for all interaction roles the user might take. Methodologies from other fields are adapted and new ones have to be developed. Previously, mostly theories and methods from psychology (e.g. [10], [11]) and human-computer interaction (HCI) have been applied to HRI research.

The aim in HCI and usability research in general is to evaluate systems with potential users in realistic conditions including environments and tasks [12], [13]. Several approaches in this direction have been taken for robotics (e.g. [14], [15]), though depending on the tasks of the robot and the context it is used in, it is often difficult to conduct live user studies. The effort to run full scale HRI trials can be large and the number of participants in HRI live user studies is usually relatively small. Another issue in live trials with a robots is that systems might show different behaviours with different users, which causes concerns regarding comparability of the data.

Video trials provide a complementary method to conduct studies with many participants and increased comparability. The authors in [6], [16] sum up the main advantages of video based HRI trials as follows: “1) reach larger numbers of subjects as they are quicker to administer, 2) easily incorporate subjects’ ideas and views into later video trials simply by recording extra or replacement scenes into the video based scenarios, 3) carry out trials exposing groups of subjects to an HRI scenario simultaneously, 4) prototype proposed live trial scenarios to avoid wasted effort and test initial assumptions, 5) allow greater control for standardised methodologies (i.e. exactly the same robot
behaviours, exact trial instructions etc.” ([16], p. 1f.).

Some major reasons lead us to apply the method in the present study. Firstly, our focus was on the evaluation of the interaction and not of technical components. Video studies do not allow for a technical evaluation of the system but, nevertheless, are suitable to research user experience. Secondly, the HRI trial can be conducted at different places (in this case Bielefeld University, Germany; University of Hertfordshire, Great Britain) with many participants at a time (e.g. in a university course). The robot effectively can be brought to the subjects, something which would have been difficult to organize in live trials in different countries. In contrast, in a video study all subjects judge the same robot behaviour and the language can be dubbed. Thus, comparability between groups is very high.

III. ROBOT SYSTEM AND SCENARIO

The robot used for the trials is called BIRON (see Fig. 1). BIRON is based on a Pioneer PeopleBot platform. A Sony EVI-D31 pan-tilt colour camera is mounted on top of the robot at a height of 142 cm to acquire images of the upper body part of humans interacting with the robot and to focus referenced objects. An additional camera is used to capture hand movements in order to recognize deictic references. A pair of AKG far-field microphones is located right below the touch screen display at a height of approximately 107 cm. They enable BIRON to localize speakers. Finally, a SICK laser range finder mounted at the front at a height of 30 cm measures distances within a scene to detect pairs of legs and to navigate.

![Fig. 1 BIRON (Bielefeld RObot CompanionN)](image_url)

The development of the robot BIRON is framed by a home tour scenario, which envisions household robots able to adjust to new environments like a user’s home. The environments have to be explored together with the customer who probably is a rather inexperienced user. She has to teach important objects and places to the robot.

Therefore, capabilities a home tour robot must exhibit for natural interaction comprise understanding of spoken utterances, co-verbal deictic reference [17], verbal output, referential feedback, and person attention and following [18].

IV. STUDY

The study presented here focused on two different robot behaviours (introvert and extrovert). Previous experiments have shown that robot personality has a major influence on HRI (e.g. [19], [20], [21], [22], [23]). It is important to keep this in mind, especially in the scenario described here. Once the robot enters the home of the person its personality becomes even more important because nobody wants to live with a robot she does not like. We therefore aim at developing a range of behaviours that allow the system to adapt to the users’ preferences. In the related work cited above, perception of personality was usually influenced by changing of the robot appearance. As [24] found for HRI and [25] for virtual agents, speech might influence human-machine interaction even more than appearance. We therefore developed two different interactive behaviours (labelled here, extrovert and introvert) based on an analysis of the verbal interaction. Moreover, we compared the effects of different movement patterns of the robot.

According to [26], extrovert personalities are described as sociable, friendly, talkative and outgoing. Introverts are quite introspective, and prefer to be with small groups of people. We tried to model these behaviours in the verbal behaviour of the robot and in the way it follows a person when entering a room. Details are described in section IV. A.

The main research questions addressed in this paper are:

- Do subjects recognize differences between the two robot behaviours (extrovert and introvert)?
- Which of the behaviours do subjects prefer?
- Is the robot displaying extrovert behaviour rated as being more friendly, intelligent and/or polite than the one displaying introvert behaviour?

In this paper we do not focus on how the robot behaviours are rated by people with different personalities because we first wanted to verify whether the robot behaviours were perceived as being distinct from each other. To test this, subjects were divided into three groups, the first two watching one robot behaviour (I: introvert or E: extrovert), the last one judging both (B). Three groups were necessary to test in-group as well as inter-group differences. All participants of a group (e.g. a course) watched the videos together. Fig. 2 displays the experimental procedure for each group. The whole experiment took about 25 to 30 minutes for the short
conditions (I and E), and 35 to 40 minutes with the longer one (B). The development of the videos and questionnaires is described in the following section.

A. Videos

The main aim of the study was to test an appropriate task in an ecologically valid environment. Therefore, the system was set up in a real apartment (Fig. 3). The video was composed of three parts. Video 1 showed the introduction to the scenario with the robot being delivered and assembled by a mechanic. The video provides additional information about the scenario of a domestic robot which can easily be purchased, set up and employed by inexperienced users. While this video was the same for all conditions, two different robot behaviours were recorded for most of the home tour (Videos 1I & 1E and 2I & 2E).

Videos 1I and 1E presented a user (enacted by a professional actress) greeting the robot and showing it objects in the living room with the robot displaying different verbal behaviours. Video 2 was identical for all conditions and presented the robot on its way from the living room to the dining room. Videos 2I and 2E showed the user guiding the robot into the kitchen, again displaying two different behaviours. The final video (Video 3) was identical for all groups, and showed BIRON driving back to the living room autonomously. Different perspectives (first person view, third person view – see Fig. 4) were included in all videos, as recommended by [16] to facilitate the viewer’s comprehension.

The two robot behaviours consisted of different verbal and movement interaction patterns. Robot behaviour I was intended to be introverted and was designed to be less proactive. The robot in this condition waited until it was addressed by the user before talking. Apart from that, the robot talked little and used brief sentences which shortened the interaction significantly (Video 1I). When the user guided the system through a door into the kitchen it needed to be steered directly by commands (Video 2I).

Robot behaviour E was rather extroverted. When the actress entered the living room, the robot addressed her instead of waiting for her to start the conversation. Moreover, the extrovert BIRON was more talkative. The robot uttered longer sentences which were also more elaborate (Video 1E). In this condition, the robot entered the kitchen autonomously. It simply followed the user instead of waiting for instructions (Video 2E). The following example illustrates the difference between the extrovert and introvert verbal behaviour:

Robot: Hello.

Extrovert (E): Robot: Hello. My name is BIRON.  
What’s your name?  
User: I’m Tina.  
Robot: Nice to meet you Tina.

B. Questionnaires

Participants filled in questionnaires to rate the interactions. Before they watched the videos, all groups received a first questionnaire, which included questions about their age, course of study, and gender and they were asked to rate their experience with computers and robots. They indicated which robots they knew out of a list of 10. After this initial questionnaire the participants watched the videos.

People rating only one robot behaviour (condition I and E) watched all the videos of their condition at once, apart from the autonomous return of the robot to the living room (Video 3). Before watching this final video, they answered
the second questionnaire. Subjects rating both behaviours (condition B) watched the interaction in the living room, answered the second questionnaire, watched the guiding to the kitchen and the interaction in the kitchen, answered the third questionnaire, and then finally watched the robot return to the living room. The questionnaires for this group contained the same items as the ones for the other two conditions. But participants answered a set of questions for each robot behaviour. Sequence effects in condition B (both behaviours) cannot be excluded since the videos were only shown to one group in each country. Thus, counterbalancing was not possible.

V. PRETEST

A pre-test was run to identify problems in the design of the study, the questionnaires, and the videos. It was conducted in German with 54 students in three different courses. Students were divided into three groups where all three conditions were tested. The pre-test brought some insights that helped us to improve the videos and the questionnaires. An advantage identified by [16], is that single video scenes can easily be changed or replaced. After the pre-test, this was advantageous regarding shortening the overall length of the video.

Participants in the pre-test watched the robot walk back to the living room before they filled in the final questionnaire. The robot travelling back autonomously to a room previously learned is an intelligent behaviour. This turned out to overshadow the differences between the two robot behaviours. We therefore decided to have participants fill in the second questionnaire before showing the concluding part of the video.

The first version of the questionnaire contained several open questions (Which robots do you know?; Name adjectives to describe the robot.). With the help of the pre-test we replaced these questions by scales to save time and to get easily comparable answers. In the pre-test subjects listed many robots they knew. Out of these answers we chose the 8 most frequently named robots (Aibo, Kismet, mars explorer, Asimo, soccer robot, Lego Mindstorms, Roomba, R2D2) plus BIRON and “service robot for the home” to explore whether people were familiar with the domain studied in these trials.

We also analyzed the adjectives people named in the pre-test to describe the robot behaviour. Groups containing words with synonymous meaning were built. Afterwards we chose the word which best described each group and adequate opposites. In the new questionnaire subjects had to rate 14 adjective pairs on a 5-point scale which again increased comparability between subjects and decreased the time to answer the questionnaire. The scale consisted of adjectives which were chosen as appropriate to divide between the two behaviours tested (active, passive; interested, indifferent; talkative, quiet) and others that might result from the perception of different robot personalities (intelligent, stupid; predictable, unpredictable; consistent, inconsistent, fast, slow; polite, impolite; friendly, unfriendly; obedient, disobedient; diversified, boring; attentive, inattentive). Some other terms investigated the general usefulness of the robot (useful, useless; practical, impractical).

VI. RESULTS

The results presented here include data acquired in a study with 200 participants in Germany (109) and Great Britain (91). All were assigned to one of the controlled experimental conditions (I (introvert)=62; E (extrovert)=72; B (both)=66). Their mean age was 23.95 years, 108 were male, 92 female. All German participants were students, whereas in GB 10 people belonged to the academic staff. 46.5% had a background in computer science (Germany: 30%, GB: 66%). The rest came from other disciplines (linguistics, German studies, media science, psychology, business, and health communication).

All participants had some experience working with computers (mean=3.97 on a scale of 1 (no experience at all) to 5 (a lot of experience)). However, most had little experience of interacting with robots (mean=1.65 on a scale of 1 (no experience at all) to 5 (a lot of experience)). Nevertheless, the majority indicated they knew some robots (mean=3.94 out of 10; min=0, max=10, sd=2.8), the best-known being: R2D2 (66.5%), Aibo (62.5%), mars explorer (49.5%), soccer robot (46%), and Asimo (45%). Only 14.5% knew BIRON.

Firstly, we analyzed the questions “How much do you like the robot?” and “How satisfied are you with the robot’s behaviour?” to find out whether subjects actually noticed a difference between the robot behaviours and if one was preferred. Table 1 presents participants’ ratings. No intercultural differences were found in this study which supports the assumption that videos with dubbed language can be shown in various countries. Anyhow, it has to be kept in mind that the sample was quite homogenous and both countries were Western European.

Table 1 illustrates that participants showed a significant preference for the extrovert robot behaviour (E). Both questions (How much do you like the robot? and How satisfied are you with the robot’s behaviour?) were answered in favour of behaviour E. To prove the significance of the differences, for conditions I and E (one

<table>
<thead>
<tr>
<th>Table 1</th>
<th>LIKEABILITY AND SATISFACTION WITH ROBOT BEHAVIOUR (MEAN ON A SCALE OF 1 (VERY LOW) TO 5 (VERY HIGH)) FOR CONDITIONS I, E, B (QUESTIONNAIRES 2 AND 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>both (quest. 2)</td>
</tr>
<tr>
<td>likeability</td>
<td>2.46</td>
</tr>
<tr>
<td>satisfaction</td>
<td>2.45</td>
</tr>
</tbody>
</table>
robot behaviour) a one-way ANOVA was calculated; for condition B (both robot behaviours) a T-Test for paired samples was conducted (Likeability condition I and E: F=21.278; df=1,130; p<.001; condition B: T=-8.231; df=64; p<.001; Satisfaction with robot behaviour condition I and E: F=5.917; df=1,132; p=.016; condition B: T=8.079; df=64; p<.001). This finding is supported by the fact that 95.2% of the subjects in condition B indicated they noticed a difference between robot behaviours in videos 2I and 2E.

However, this does not hold true for the second rating of group B that judged both robots after the kitchen entry scene. Even though 63.5% noticed a difference between robot behaviours I and E, the ratings of the likeability and satisfaction with the robot behaviour did not differ. Reasons are given in the following analysis of the adjective ratings of the behaviours.

For the analysis of the adjectives, again a one-way ANOVA was calculated for the groups that rated one robot behaviour (Table 2), and a T-Test for the group that rated both the extrovert and introvert behaviour (Table 3). The differences between the ratings of behaviour I and E were obvious. Firstly, behaviour E was rated as being significantly more active, talkative, and interested. This proves that the modelling of the behaviours was successful. However, results were different for the second rating of the behaviours by group B after the kitchen entry scene. The participants could not distinguish between extrovert and introvert behaviour. Neither door crossing was preferred. This result might partly be due to the fact that the door crossing scenes were insufficiently meaningful and the viewers of the videos did not recognize a difference. However, there might be strong preferences for one person following behaviour in live user studies where people actually might feel comfortable with the robot or not.

The following analysis focuses on the ratings of group I and E, and of group B after the interaction in the living room (questionnaire 2). Tables 2 and 3 show that the difference between these ratings is greater when people watched both robot behaviours, and were therefore able to compare them. This might partly be due to the fact that groups I and E rated the robot only once after the kitchen entry scene. However, the tendency of the results was the same for most items. In general, the extrovert robot behaviour was rated significantly more friendly, diversified, fast, and intelligent.

### Table 2

One-way ANOVA of ratings of Robot Behaviours in Condition I and E (mean on a scale of 1 (not at all) to 5 (very much), F-value (df=1, 132), and significance)

<table>
<thead>
<tr>
<th>item</th>
<th>mean I</th>
<th>mean E</th>
<th>F value</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>active</td>
<td>2.30</td>
<td>2.89</td>
<td>12.247</td>
<td>.001**</td>
</tr>
<tr>
<td>talkative</td>
<td>2.00</td>
<td>2.93</td>
<td>26.145</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>interested</td>
<td>2.87</td>
<td>3.26</td>
<td>5.358</td>
<td>.022*</td>
</tr>
<tr>
<td>attentive</td>
<td>3.54</td>
<td>3.59</td>
<td>.105</td>
<td>.747</td>
</tr>
<tr>
<td>fast</td>
<td>1.61</td>
<td>2.00</td>
<td>6.813</td>
<td>.010*</td>
</tr>
<tr>
<td>consistent</td>
<td>3.22</td>
<td>3.31</td>
<td>.361</td>
<td>.549</td>
</tr>
<tr>
<td>predictable</td>
<td>3.35</td>
<td>3.44</td>
<td>.256</td>
<td>.614</td>
</tr>
<tr>
<td>polite</td>
<td>4.05</td>
<td>4.31</td>
<td>3.149</td>
<td>.078</td>
</tr>
<tr>
<td>friendly</td>
<td>3.56</td>
<td>4.03</td>
<td>9.218</td>
<td>.003**</td>
</tr>
<tr>
<td>obedient</td>
<td>4.16</td>
<td>4.31</td>
<td>.894</td>
<td>.346</td>
</tr>
<tr>
<td>diversified</td>
<td>2.12</td>
<td>2.69</td>
<td>9.218</td>
<td>.003**</td>
</tr>
<tr>
<td>intelligent</td>
<td>2.98</td>
<td>3.34</td>
<td>4.433</td>
<td>.037*</td>
</tr>
<tr>
<td>practical</td>
<td>2.10</td>
<td>2.27</td>
<td>1.026</td>
<td>.313</td>
</tr>
<tr>
<td>useful</td>
<td>2.18</td>
<td>2.13</td>
<td>.087</td>
<td>.768</td>
</tr>
</tbody>
</table>

### Table 3

T-Test for paired samples for rating of Robot Behaviours in Condition B (mean on a scale of 1 (not at all) to 5 (very much), T (df=65), and significance for ratings questionnaire 2 (questionnaire 3))

<table>
<thead>
<tr>
<th>item</th>
<th>mean I</th>
<th>mean E</th>
<th>T</th>
<th>signific. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>active</td>
<td>2.23</td>
<td>3.85</td>
<td>-12.734</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>talkative</td>
<td>1.97</td>
<td>4.02</td>
<td>-15.086</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>interested</td>
<td>2.41</td>
<td>3.89</td>
<td>-10.841</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>attentive</td>
<td>2.68</td>
<td>3.50</td>
<td>-7.083</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>fast</td>
<td>2.02</td>
<td>2.58</td>
<td>-4.511</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>consistent</td>
<td>3.30</td>
<td>3.39</td>
<td>-7.600</td>
<td>.450</td>
</tr>
<tr>
<td>predictable</td>
<td>3.42</td>
<td>3.06</td>
<td>2.168</td>
<td>.034*</td>
</tr>
<tr>
<td>polite</td>
<td>2.98</td>
<td>4.12</td>
<td>-7.855</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>friendly</td>
<td>2.88</td>
<td>4.03</td>
<td>-9.114</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>obedient</td>
<td>3.55</td>
<td>3.56</td>
<td>-1.36</td>
<td>.892</td>
</tr>
<tr>
<td>diversified</td>
<td>1.88</td>
<td>3.02</td>
<td>-8.550</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>intelligent</td>
<td>2.65</td>
<td>3.55</td>
<td>-7.421</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>practical</td>
<td>2.21</td>
<td>2.35</td>
<td>-1.732</td>
<td>.088</td>
</tr>
<tr>
<td>useful</td>
<td>2.21</td>
<td>2.45</td>
<td>-2.248</td>
<td>.026*</td>
</tr>
</tbody>
</table>
Politeness was rated very high in both conditions (I and E). Only the direct comparison of group B shows that the extrovert behaviour was judged as being significantly more polite. Altogether, the results indicate that the verbal behaviour of the robot is a powerful means to model robot personality traits.

Even though a clear preference for robot behaviour E was found, the behaviour only had a small effect on the perceived usefulness of the system. In all the conditions, people did not rate the robot as being very useful or practical (see Tables 2 and 3). Reasons for this are, firstly, that BIRON did not perform any manipulative tasks in the video because the study focused on more general behaviour. Secondly, the robot used has no kind of manipulator to actually provide services in the household, such as picking up glasses for example.

VII. CONCLUSION

The video-based study methodology has the main advantage of reaching many participants (200) in geographically distant places in a very short time. Participants viewed videos of BIRON and rated the two different behaviours exhibited by the robot (extrovert, introvert). They noticed the differences between the behaviours and preferred the extrovert robot. Traits like intelligence, interest, friendliness, and diversity were more strongly associated with extrovert behaviour, which is also true in human-human interaction.

These attributions were found mainly to be a result of the dialogue design, because they did not hold true for a door crossing behaviour. However, this finding might be due to restrictions of the video based HRI method, and might change in live user studies. Nevertheless, the study gives helpful insights into users’ preferences, which guide the current system design and implementation. It can be considered a powerful supplement to interactive user studies in realistic settings with working prototypes.

REFERENCES