Practical and Methodological Challenges in Designing and Conducting Human-Robot Interaction Studies

Michael L. Walters, Sarah Woods, Kheng Lee Koay, Kerstin Dautenhahn
{m.l.walters, S.N.Woods, K.L.Koay, K.Dautenhahn}@herts.ac.uk
Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire
College Lane, Hatfield, AL10 9AB

Abstract

Human-robot interaction is a rapidly growing research area which more and more roboticists and computer scientists are moving into. Publications on work resulting from such studies rarely consider in detail the practical and methodological problems encountered. This paper aims to highlight and critically discuss such problems involved in conducting human-robot interaction studies. We provide some examples by discussing our experiences of running two trials that involved humans and robots physically interacting in a common space. Our discussion emphasises the need to take safety requirements into account, and minimise the risk of physical harm to human subjects. Ethical considerations are considered, which are often within a formal or legal framework depending on the host country or institution. We also discuss future improvements for features of our trials and make suggestions as to how to overcome the challenges we encountered. We hope that the lessons learnt will be used to improve future human-robot interaction trials.

1 Introduction

In the course of our research for the COGNIRON Project [2005], we are primarily interested in the research area of Human-Robot Interaction (HRI), in particular with regard to socially interactive robots. An excellent overview of socially interactive robots (robots designed to interact with humans in a social way) is provided in Fong et al. [2003]. As we are primarily studying the human perspective of human-robot interaction, human scaled robots in live trials within a human orientated environment were required.

Fig. 1: Children playing games with the University of Hertfordshire PeopleBot™ at the London Science Museum event in October 2004.

Other researchers that have conducted similar human centred trials with human sized robots include Dario et al. [2001], Severinson-Eklundh et al. [2003], Kanda et al. [2004] and Hinds et al. [2004]. To date, we have conducted two human-robot trials with human scaled PeopleBot™ robots. One trial involved a single robot interacting with groups of children in a game scenario. The trials took advantage of a software evaluation event at the University of Hertfordshire, hosted by the Virtual ICT Empathic Characters (VICTEC) project [VICTEC, 2003]. The other trial involved individual adults interacting with a robot in various contexts and situations, within a simulated domestic (living-room) environment. We have also participated in other displays and demonstrations which have involved robots interacting in the same physical space as one or more humans. In particular, we successfully ran interactive games for groups of up to 40 children at a time, at a major public event at the Science Museum in London [BBC Science News, 2004]. The PeopleBot™ robots have also been demonstrated on several occasions during open days at the University of Hertfordshire. This paper will present some of the methods we have developed and critically discuss the various trials and events we have been involved with to date.

2 Planning, Legal and Safety

Before running a trial involving humans and robots physically interacting, certain legal and ethical issues must be satisfied. At this stage it is good practice, and in the UK a legal requirement under the Management of Health and Safety at Work Regulations 1999, to carry out a risk assessment for all work activities involving employees or members of the public [Crown Copyright, 2003]. These first activities are considered here.

2.1 Legal and Ethics Approval

Many institutions, including the University of Hertfordshire [UPR AS/A/2, 2004], require that an Ethics Committee must give approval for all experiments and trials involving human subjects. Usually, this approval is gained by submitting a (written) description of the trials or experiments to be performed to the committee. The Ethics committee will then consider the proposal, and may modify, request further clarification, ask for a substantial rewrite, or even reject the proposal outright on ethical grounds. In general, the Ethics committee will make possible objections on the following grounds:

Privacy – If video, photographic or records of personal details of the subjects are being made and kept, the committee will be concerned that proper informed consent is given by subjects, any personal records are securely stored and will not be misused in any way. If personal data is to be held on a computer database, then the legal requirements of the Privacy and Electronic Communications Regulations [Crown Copyright, 2003] must be adhered to. If any public use of the video or photographs is to be made for conferences or publicity purposes, then participants must give explicit permission.

Protection of minors and vulnerable adults – In the UK it is a legal requirement (Protection of Children and Vulnerable Adults Order, 2003) that anyone who works with children or vulnerable adults must have their criminal record checked. In the UK, anyone under 18 is classed as a child in this context, and the term vulnerable adult includes the infirm or elderly in a care situation. Regulations in many other countries in Europe are less strict, but if experiments or trials are planned to involve children or vulnerable adults, then any legal implications or requirements must be considered. For example, not gaining the appropriate checking of criminal records could lead to a situation where subjects who are keen to participate in a study need to be turned down. Given the general problem in recruiting a sufficiently large sample of human subjects, this could potentially cause problems.

Mental or emotional stress and humiliation – The trials should not give rise to undue mental or emotional stress, with possible long-term repercussions. Where an experimental situation is actually designed to put a subject under stress intentionally, it may not be possible to avoid stressing the subject. The Ethics Committee will want to be satisfied that if any mental or emotional stress is suffered by subjects, it is justified and that no after effects will be suffered by subjects. In our own studies we were interested in how subjects ‘spontaneously’, or ‘naturally’ behaved towards robots, so we had to carefully design the scenarios in order to be on the one hand controlled enough to be scientifically valuable, but on the other hand open enough to allow for relaxed human-robot interactions. It is advised to include a statement in the consent form which points out that the subject can interrupt and leave at any stage during the trial for whatever reasons, if he or she wishes to.

Physical harm – Practically all experiments that involve humans moving will involve some degree of risk. Therefore, any human-robot trials or experiments will pose some physical risk for the subjects. The Ethics Committee will want to be satisfied that the proposal has considered any potential physical risks involved. The subjects’ safety is covered in more detail below.

2.2 Safety

Robot-Human Collision Risk - For trials involving humans and robots, the obvious immediate risk is the robot colliding with a human subject, or vice versa. The robots that we used in our trials are specifically marketed for the purpose of human-robot interaction studies. In order to alleviate the risk of the robot colliding with human subjects, two strategies were adopted:

Overriding anti-collision behaviour – The PeopleBot™ robots we use can have several behaviours running at the same time as any top-level program. This is a natural consequence of the PeopleBot™ operating system which follows the principles of the subsumption architecture expounded originally by Rodney Brooks [1991]. Many other commercially available robot systems have similar programming facilities. We always had basic collision avoidance behaviours running at a higher priority than any task level program. This means that no matter what the task level program commands the robot to do, if a collision with an object is imminent, the underlying anti-collision behaviour cuts in. Depending on the form of the hazard and the particular safety behaviour implemented, the robot will either stop or turn away from the collision hazard. The lower priority task level programs include both those that provide for direct or semi-autonomous remote control by Wizard of Oz (WOZ) operators [Maulsby et al. 1998] and also fully autonomous programs. We have found that the sonar sensors used by the PeopleBot™ are very sensitive to the presence of humans. However, some common household objects, especially low coffee tables, are not so readily picked up by the sensors. By judicious placing of objects that are readily sensed, such as boxes, footballs, cushions etc, it is possible to create a trial environment where it is literally impossible for the robot to collide with any object. For example, we adopted this strategy to avoid the robot bumping into the table where the person was sitting (see fig.2).

Fig. 2: A subject sitting at the desk, showing a box placed under the table to create a target for the robot's sonar sensors.

Monitoring by the WOZ operators – Even while the robot is running a fully autonomous program, a WOZ operator (see section 3.1) monitors discreetly what is happening. The robot's underlying safety behaviours include the overriding ability for the WOZ operator to stop the robot immediately by remote wireless link if it is perceived that the robot poses a risk to a human at any time. There is also a large red emergency stop button on the robot, which is hardwired, providing an independent failsafe method to stop the robot. Simply pressing the button cuts the power to all the robot's motors. This is simple enough for non experts to operate, and will work even if the robots control software crashes or fails to respond. Anyone who is physically close enough (i.e. in perceived danger) to the robot can access the button.

In our trials, only during the software development process of the program has it been necessary for WOZ operator or others to initiate a stop; mainly to avoid the robot damaging itself rather than actually posing a threat to humans in the vicinity. During our human-robot interaction trials, the underlying safety behaviour has proved to be both robust and reliable in detecting and avoiding collisions with both children and adults. The actual robot programs have been heavily tested in the physical situations for all the trials we have run. This is necessary as knowing how the robot will respond in all physical circumstances is critical for the safety of the participants in any trial.

For the risk case of a human colliding with the robot, there is little action that can be taken by the robot to avoid a human. The robot moves and reacts relatively slowly, compared to the speeds achievable by a human. Therefore, it is up to the human to avoid colliding with the robot. Luckily, most humans are experts at avoiding collisions and we have found that none of our subjects has actually collided with the robot. In some of the trials with children it has been necessary to advise the children to be gentle or to move more carefully or slowly when near to the robot. We found that children will mostly take notice if the robot actually issues these warnings using the robot's own speech synthesis system.

Other Possible Risks to Participants - Our robot was fitted with a lifting arm, which had a small probability of causing injury to humans. The arm itself was made of coloured cardboard made to look solid, so it looked more dangerous than it actually was. Our main concern about the arm was if the 'finger' was accidentally pointed into a human's face or eyes. This risk was minimised by keeping the arm well below face level even when lifted. Other possible risks to participants that must be considered are those that would be present in any domestic, work or experimental situation. These include things such as irregular or loose floor coverings, trailing cables, objects with sharp or protruding edges and corners, risk of tripping or slipping, etc.

In our trial involving children, small prizes were given during and at the end of each session. We were advised against providing food (i.e., sweets) as prizes, as some children may have had allergies or diabetes which could be aggravated by unplanned food intake. We also never left subjects alone with a robot without monitoring the situation.

3 Experimental Implementation

When running a human-robot interaction trial, the question that must be addressed is how to implement the proposed robot functions and behaviour. There are two main methods for developing suitable robot features, functions and behaviour for trials where we are primarily interested in the human-centred perspective towards the robot or its function.

3.1 Wizard of Oz Methods

It is usually relatively quick to create a scenario and run the robot under direct WOZ operator control. This is a technique that is widely used in HRI studies as it provides a very flexible way to implement complex robot behaviour within a quick time-scale (Robins et al. 2004 and Green et al. 2004). The main advantage is that it saves considerable time over programming a robot to carry out complex interactions fully autonomously. However, we have found that it is very tiring for the WOZ operators to control every aspect of the robot's behaviour, especially in multi-modal interactions and scenarios. It usually requires two operators, one for controlling movement and one for speech, in order to maintain reasonable response times during a trial. It is also difficult to maintain consistency between individual trial sessions. Practise effects are apparent as the operators become better at controlling the robot at the particular task scenario through the course of a series of trials. Practise effects can be minimised by thoroughly piloting the proposed scenario before carrying out 'live' trials.

Fig. 3: The Wizard of Oz operators and control room area for human-robot interaction trials at the University of Hertfordshire in 2004.

3.2 Autonomous Robot Control

The other robot control method is to pre-program the robot to run all functions autonomously. Obviously this method overcomes the problems of operator tiredness and consistency, but implementing complex autonomous behaviour is very time-consuming. However, if trials are testing complex human-robot social behaviours, or implementing desired future robot capabilities, it will not be technically feasible at present to program a robot to act fully autonomously. In accordance with the COGNIRON project aims, we are studying scenarios that go "beyond robotics". For this we have to project into the future in assuming a robot companion already exists that can serve as a useful assistant for a variety of tasks in people's homes. Realistically, such a robot does not yet exist.

The PeopleBot^{TM} robots have a sophisticated behaviour based programming API called ARIA [ActivMedia Robotics, 2005]. This provides facilities to develop task control programs, which can be integrated into the ARIA control system. The actual task control program can be assigned a priority, which is lower than the previously mentioned safety behaviours (see section 2.2). Therefore, fundamental safety and survival behaviour, such as collision avoidance, emergency stop etc. will always take precedence over the actual task commands.

In practice, we have found that a mixture of autonomous behaviours and functions, and direct WOZ control provides the most effective means of generating the desired robot's part of the HRI. The basic technique is to pre-program the robot's movements, behaviours or sequences of movements, as individual sequences, gestures or actions that can be initiated by the WOZ operator. In this way the WOZ operator is able to exercise judgement in initiating an appropriate action for a particular situation, but is not concerned with the minute details of carrying that action out. The operator then is able to monitor the action for potential hazard situations and either stop the robot or switch to a more appropriate behaviour. Because the robot is actually generating the individual movements and actions autonomously, better consistency is ensured. Also, the temporal behaviour of a robot under WOZ or autonomous control is likely to differ significantly, so whenever possible and safe, autonomous behaviour is advantageous over remote-controlled behaviour.

Robot program development & *pilot studies*- When developing robot programs, which will be used to implement a HRI trial scenario, it is important to allow enough time to thoroughly practise the programs and scenarios thoroughly before the actual trials take place. Pilot studies should be conducted with a variety of humans, as it is easy for the programmer or operator to make implicit or explicit assumptions about the way that humans will behave in response to a given trial situation. Of course, humans all exhibit unique behaviour and can do unexpected things which may cause the robot program to fail.

The first trials we ran involved interactive game sessions with groups of children. These required the children to play two short games with the robot, a *Rotation game* and a *Wander game*. The game programs ran mostly autonomously, except for starting the respective game programs, and also at the end of each round where a winning child was selected manually by remote control. When developing the interactive game programs for the Science Museum visit, the games ran totally autonomously for the whole of each game session. The Science museum game program was more complex than the previous child group games programs as sensor interpretation was involved. However, because the Science Museum robot game program was fully autonomous, the pre-testing phase had to be much longer. The extra time was needed to empirically find out optimum action and response timings and durations, sensor levels and cues, and refining the program so that it worked properly with all the human test subjects.

For the single adult HRI trials, there were time limitations on setting up and implementing the experiment. The robot behaviour was implemented almost entirely by direct WOZ control (with overriding safety behaviour active). There was also limited time available for practicing the scenarios, which were to be implemented for the study. The only autonomous behaviours used for this study were the wandering behaviour, used for acclimatising the subject to the robot's presence, and the arm lift height, which was used to set the arm to the correct height for picking up special pallets which contained items that would be fetched by the robot at various times during the trial (fig 4).

Fig. 4: The robot, fitted with a hook-like end-effector, was able to fetch small items in special pallets.

The WOZ operators were out of direct sight of robot and subject, and observed the scenarios via network video cameras placed around the room. The images from these were delayed by approximately 0.5 sec. There was also a direct, but restricted, video view from the robot camera which did not have any discernable delay. These factors made providing timely responses (comparable to human responses) to the subject very difficult for the WOZ operators. However, it can be argued that, in the near future at least, this is likely to be true of all robots, and this was a realistic simulation of likely future robot performance.

4 Video Recording

It is desirable to make a complete video record of the trials. Video footage is one of the primary means of gaining results for later analysis and validation of results. They can be used to validate data obtained by other means, e.g. from direct measurement, questionnaire responses, or recorded sensor data. Good video footage can provide time stamped data that can be used, processed and compared with future studies. However, in addition to the obvious advantages of video data, there are some drawbacks that researchers should be aware of at the outset of the design phase. Analysing video footage is an extremely time consuming process and requires thorough training in the application of the scoring procedures, which can be complex. Observations made from video footage are subjective and the observer may portray their own perceptions and attitudes into the date. For this reason, it is essential that a full reliability analysis of video data is carried out involving independent rating and coding by observers who were not involved with the study, and did not meet any of the participants.

4.1 Video camera types

We used two types of video cameras for recording our trials; tripod mounted DV camcorders, and network cameras. The DV camcorders record onto mini DV tape, which must then be downloaded onto a computer hard disk before further analysis can be performed. The network cameras have the advantage that they record directly to a computer's hard disk, so there is no tedious downloading later on. They do require some synchronising, converting and combining, but this can be done automatically in batches overnight. We have found that the DV Camcorders provide a better quality picture than the network cameras, with a synchronised soundtrack. While high quality video may not be strictly necessary for analysis purposes, it does allow high quality still pictures to frame-grabbed from the video recordings, which are invaluable for later writing up, papers and reports. It is also easy to create short videos to incorporate into presentations and demonstrations using standard video editing software.

It is advisable to use at least two camera systems for recording trials or experiments. If one camera fails, than there will another stream of video data available. It should be borne in mind that if a network camera fails, it may also lead to all the network cameras being bought down. Therefore, at least one camera should be a freestanding camcorder type, which stores the video data on (mini DV) tape.

Note, a similar backup strategy is also advisable as far as the robotic platforms are concerned. In our case, we had a second PeopleBot™ in place, in the event that one robot broke down. Having only one robot available for the trials is very risky, since it could mean that a trial had to be abandoned if a robot fails. Re-recruiting subjects and properly preparing the experimental room is a very time-consuming activity, unless a permanent setup is available. This was not the case in our trials, where rooms were only temporarily available for a given and fixed duration (two weeks for the study involving children, 2 months for the adult study). Afterwards the setups had to be disassembled and the rooms had to be transformed back into seminar or conference rooms. This also meant that any phases of the trials could not be repeated. Therefore, it was essential to get it right first time despite the limited preparation time. This is a situation common to a University environment with central room allocations and usually few permanent large laboratory spaces suitable for studies with large human-sized robots.

4.2 Camera Placement

The placement of the cameras should be such that the whole trial area is covered by one or two views. For our first trial, we used two cameras placed in opposite corners of the room, both facing towards the centre of the room. As a result we recorded two views of the centre of the room, but missed out on what was happening at the edges of the room. A better way to position the cameras would have been to point the cameras to the right (or left) of room centre, with only a small view overlap in the centre of the room. This way, the two views also include the outer edges of the room. (See fig.5)

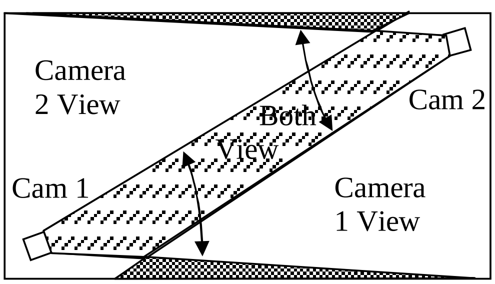

Fig. 5: Diagram showing correct placement of cameras to maximise coverage of room.

It is best to use two cameras to cover the entire area as shown in the diagram, with additional cameras to obtain detailed views of specific areas of interest. For example; when it is known that subjects will have to sit at a certain desk, which is in a fixed position, it is worth setting up an individual camera just to record that position in detail. Setting the correct height of the cameras is important to obtain a good view of the subjects face.

4.3 Distance Measurements

One main aspect of our trials was focused on examining the spatial distances between the robot and human subjects. Video images can be useful in estimating these distances. In both our child group and single adult trials, markings were made on the floor with masking tape to provide a method to estimate the position of the robot and the human subjects within the trial areas. However, these markings were visible to the subjects, and may possibly have influenced the positioning of the human subject during the course of parts of the experimental scenarios.

In the context of a study described by Green et al. [2004], a method was used that involved overlaying a grid of 0.5m squares onto still images of the floor of their trial area for individual frames from their video recordings. This method would allow the positions of the robot and subject to be estimated with a high degree of accuracy if it can be adapted for live or recorded video data. It would provide a semi-transparent grid metric overlaid onto the floor of the live or recorded video from the cameras. The possibility of visible floor markings affecting the positions taken by the subject would not happen. For future trials we will want to use such a 'virtual grid' on the floor of the recorded video data. We are currently evaluating suitable video editing software.

5 Subject's Comfort Level

For the adult trials, we experimented with a method of monitoring how comfortable the subject was while the trials were actually running. We developed a hand held comfort level monitoring device (developed by the first author) which consisted of a small box that could be easily held in one hand (see fig. 6). On one edge of the box was a slider control, which could be moved by using either a thumb or finger of the hand holding the device. The slider scale was marked with a happy face, to indicate the subject was comfortable with the robot's behaviour, and a sad face, to indicate discomfort with the robot's behaviour.

Fig. 6: Photograph of Hand Held Comfort Level Monitoring Device

The device used a 2.4GHz radio signal data link to send numbers representing the slider position to a

PC mounted receiver, which recorded the slider position approximately 10 times per second.. The data was time stamped and saved in a file for later synchronisation and analysis in conjunction with the video material. The data downloaded from the hand held subject comfort level device was saved and plotted on a series of charts. However, unexpectedly, the raw data was heavily corrupted by static from the network cameras used to make video recordings of the session. It has been possible to digitally clean up and recover a useful set of data. A sample of the raw data and the cleaned up version is shown in the figs. 7 and 8.

Many of the comfort level movements correspond to video sequences where the subject can be seen moving the slider on the comfort level device. This confirmed that the filtered files were producing a reliable indication of the comfort level perceived by the subject. For future trials, it is intended to incorporate error checking and data verification into the RF data transfer link to the recording PC in order to reduce problems with static.

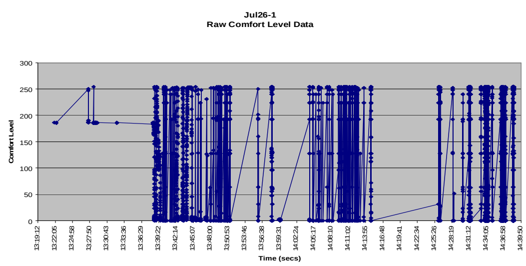

Fig. 5: Raw Data as Received from Handheld Comfort Level Monitoring Device

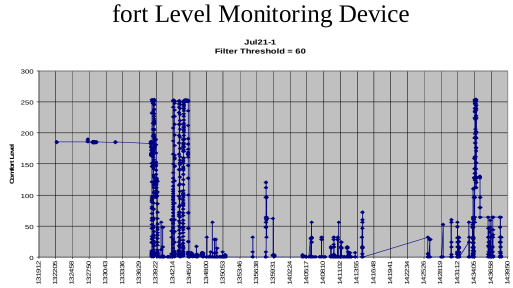

Fig. 6: Digitally Filtered Data from Handheld Comfort Level Monitoring Device

6 Questionnaires

For both interactive trials, subjects were asked to complete questionnaires. For the child-robot interactions only five minutes at the beginning, and five minutes at the end of each session were available. Due to limited time, only basic information was obtained, such as gender, age, approval of computers and robots, and how they liked the interactive session. For the adult study, the questionnaires were much more comprehensive. The time taken for the session typically ranged from 40 minutes to 1 hour. Up to half the time was spent completing questionnaires. The questionnaires covered the subjects' personality traits, demographics, technical experience, opinions towards a future robot companion, how they felt about the two contrasting robot 'personalities' exhibited by the robot during the interaction scenarios, what they liked or disliked about the robot interactions, and how it could be improved, etc.

6.1 Questionnaire Design

Questionnaire design requires training and experience. There are a number of different considerations researchers should take into account before embarking on designing a questionnaire. Firstly, the notion of whether a questionnaire is the best form of data collection should be addressed. For instance, in some situations an interview format might be preferred (e.g. if conducting robot interactions studies with young children that have low reading abilities). A questionnaire is usually completed by the participant alone. This does not allow the researcher to probe for further information they feel may be relevant to the experiment or verify participant responses. However, the advantage of using questionnaires is that they are usually fast to administer, and can be completed confidentially by the participant.

The development of a questionnaire goes through a series of different cycles. Questions that should be considered are:

- Is the questionnaire I am going to use a valid measure (i.e. does it measure what I really want it to measure)
- Is it reliable (i.e. do I get the same pattern of findings if the questionnaire is administered a few weeks later?),
- Have I used value-laden or suggestive questioning (e.g. "Do you think this robot is humanlike?"), compared to neutrally phrased questions (e.g. "What kind of appearance do you think this robot has?")?,
- Do I want to use a highly structured questionnaire or a semi-structured questionnaire, for example where subjects can express their attitudes towards a particular aspect of the robot interaction in more detail?

Some questionnaires are easier to design than others. For example, a questionnaire that enquires about subject demographics must include items that enquire about age and gender. However, even when considering something as simple as age, the researcher must decide whether to use age categories or simply get the subject to write their age in.

The complexity of questionnaire design occurs when a new research domain is being explored, and human-robot interaction is a perfect example of this. There is no such thing as a perfect questionnaire, but careful team planning and pilot testing can ensure that you have the best possible measure. To carry out a pilot test for a questionnaire, the researcher must recruit independent subjects with the same demographics that they hope to include in the real experiment. Sometimes, it is not easy to get volunteers to participate in a pilot test, but obviously the more responses you get, the more certain you can be of what necessary changes need to be made. It is good practise to carry out the pilot study with approximately 5-10 subjects although this depends on the number of conditions etc in the experiment. In addition to asking the pilot subjects to complete the trial questionnaires, it is recommended to ask them directly whether they found any aspects particularly unclear, complicated or irrelevant etc. One could also ask the subjects whether they would change anything about the overall structure or format, and whether there were important questions that you omitted.

A further issue relates to the type of data you will have to analyse. It is important at the design and pilot testing phase to consider the statistical frameworks that you wish to use, as the questions need to be asked in order to fit their requirements as well as the research goals. For example, continuous scales for questionnaire responses lead to very different analytical frameworks compared to categorical (e.g. yes/no) response formats (i.e. interval versus nominal/ordinal data). Although this process can seem time consuming at the outset, it is certainly worth it, as it is impossible to make changes while the trials are running. An error in the questionnaires could possibly invalidate one or more questions, or in the worst case, the whole questionnaire. As highlighted above, no questionnaire is perfect and we discovered this for ourselves in the adult robot-interaction study. Below we give an example of a possible problematic question and a suggested solution:

Example question

Q. Would you like the robot to approach you at a speed that is?

1) Fast
2) Slow
3) Neither fast nor slow

The above question is phrased in an unspecific way, resulting in, whatever answer is given, little quantifiable information about the preferred approach speed. Due to this lack of a reference point, in practice, most subjects are likely to choose answer 3), as most people want the robot to approach at a speed which is 'just right'. An improved way of asking the question could be:

Suggested improvement to question

Q. Did the robot approach you during the trials at a speed that you consider to be?

1) Too fast
2) Too slow
3) About right

Hopefully, results obtained from this improved question would relate a subject's preferred robot approach speed relatively to the actual speed employed by the robot in a trial. If finer graduations of preferred robot approach speeds are desired, then the trial context and situation must be more closely controlled, with multiple discrete stages, with the robot approaching at different speeds at each stage.

Questionnaire Completion - In our trials it was necessary for some of the questionnaires to be completed in the robot trial area. The subject completed the first questionnaires while the robot was wandering around the trial areas in order to acclimatise the subject to the robot's presence. The two post scenario questionnaires were also administered in the trial area, straight after the respective scenarios, while they were fresh in the subject's memory. We were not able to gain access to the trial area to turn the video cameras off during this time, as we wanted to preserve the illusion that the subject and supervisor were on their own with the robot during the trial. However, there were several other questionnaires and forms, which could have been, administered elsewhere. This would have reduced the amount of video tape used per session. Also the WOZ operators had to sit perfectly still and quiet for the duration of these questionnaires. However, a drawback of administering the questionnaires outside the experimental room is that it changes the context, and might distract the subject etc. Such factors might influence the questionnaire results. Thus, there is a difficult trade-off between savings in recording video tape and other data during the trials, and providing a 'natural' and undisturbed experimental environment.

The environmental context is an important consideration for human-robot interaction studies as questionnaire and interview responses, and observational data will vary depending on the experimental set-up. For example, it would not appear to be problematic to complete a participant demographics questionnaire in the experimental room, which in this case was the simulated living room containing the robot. However, when administering a questionnaire that relates to robot behaviour, appearance, personality or the role of future robot design, the robot and room set-up could influence subject responses. For example, in both the child and adult studies subjects completed a questionnaire at the end of the robot interaction scenarios about their perceptions towards a future robot companion. If the intention is that they consider the robot interaction and robot appearance they have just interacted with in the responses (as it was the case in our study), then this is acceptable. However, the researchers must be aware that subject experiences with the robots in the simulated living room are likely to have influenced their responses in some way.

For trials run in 2004 at the Royal Institute of Technology, Sweden, the WOZ and camera operators were in view of the subject while user trials were taking place [Green et al. 2004]. However, the focus of their study was mainly on human-robot dialogue and understanding, command and control of the robot, which may not have been affected by the presence of other people. We have found that when other people are present, then subjects will tend to interact with those other people, as well as the robot. For our single adult interactions, we wanted to observe the subjects reactions as they interacted only with the robot. Thus, while the experimenter in the adult study stayed in the same room as the subject, she deliberately *withdrew* herself from the experiment by sitting in a chair in a corner and reading a newspaper. Moreover, she did not initiate any communication or interaction with the subjects, apart from situations when she had to explain the experiment or the questionnaires to the subject, or when she had to respond to a verbal query from the subject. We opted for this approach since the study targeted a '*robot in the home'* scenario, where it would be likely that a person and robot would spend a considerable amount of time alone together in the environment.

7 Design and Methodological Considerations

At the outset of designing any study there are a number of crucial design and methodological considerations.

First, the research team must decide what the sample composition will be including, individuals, groups, children, adults, students, or strangers from the street. This is important as the interpretation of results will be influenced by the nature of the sample. For example in the current study, we observed quite distinct differences in the interaction styles between groups of children who were familiar with each other, and individual adults who were alone in the room with an experimenter who did interact in the experiment. Also, as with many other studies, the current adult sample were self-selected and were all based at the university (either as staff or students), which could result in a positive or negative bias in the results. It is very difficult to recruit completely randomised samples and there is always a certain amount of self-selection bias in all studies of this design.

Second, the environmental context should be considered, in the sense of whether a laboratory set-up is used or a more naturalistic field study. Different results are likely to emerge depending on the environment chosen. The adult human-robot interaction study involved a simulated living-room situation within a conference room at the University. Although we tried to ensure it was as realistic as possible, subjects still knew it was not a real living room and were likely to have felt monitored by the situation. Ideally it would be best to carry out future robot-human interaction studies in peoples' homes or work places in order to capture more naturalistic responses and attitudes towards the interactions. However, there are advantages for carrying out laboratory based studies as it allows the researchers greater control and manipulation of potential confounding variables. This cannot be done in the naturalistic field, so it is certainly common practise to begin new research protocols in laboratory set-ups.

Cultural differences are also important if the researchers are hoping for widespread generalisation of the findings. However, this is often impractical, highly expensive and time-consuming.

The overall design of experiments is extremely important in terms of whether between-subject groups (independent measures design) or within-subject groups (repeated measures designs) are used. There are advantages and disadvantages associated with both. Between-subject designs involve different subjects participating in different conditions, whereas within-subject designs mean that the same set of subjects take part in a series of different conditions. Between-subject designs are less susceptible to practice and fatigues effects and are useful when it is impossible for an individual to participate in all experimental conditions. Disadvantages include the expense in terms of time and effort to recruit sufficient participant numbers and insensitivity to experimental conditions. Within-subject designs are desirable when there are sensitive manipulations to experimental conditions. As long as the procedures are counterbalanced, biased data responses should be avoided.

A final consideration should be whether the researchers feel the results are informative based on information recorded at one time point. Human-robot interaction involves habituation effects of some kind and it would be highly useful for researchers to be able to follow-up the same sample of subjects over an extended period of time at regular intervals, to determine whether for example, they become more interested/less interested in the robot, more positive/negative towards the robot and so forth.

Human-robot interaction studies are still a relatively new domain of research and are likely to have a high explorative content during initial studies. It took the science of human psychology many years to build up a solid base of methods, techniques and experience, and this process is still going on at the present. The field of human-robot interactions is still in its infancy and carrying out these initial explorative studies implies that there are not likely to be any concrete hypotheses claiming to predict the direction of findings. This would be impossible at the outset of studies if there are not many previous research findings to base predictions on. The nature of exploratory studies means that there are likely to be many different research questions to be addressed and in any one study, it is simply impossible to consider all possible variables that might influence the findings. However, once exploratory studies have been conducted it should allow the researchers to direct and elucidate more concrete and refined research hypotheses for future, more highly controlled studies.

8 Summary and Conclusions

We have discussed our experiences of running two trials that involved humans and robots physically interacting, and have highlighted the problems encountered.

1. When designing and implementing a trial that involves human and robots interacting physically within the same area, the main priority is the human subject's safety. Physical risk cannot be eliminated altogether, but can be minimised to an acceptable level.
2. There are ethical considerations to be considered. Different countries have differing legal requirements, which must be complied with. The host institution may also have additional requirements, often within a formal policy.
3. Practical ways are suggested in which robots can be programmed or controlled in order to provide intrinsically safe behaviour while carrying out human-robot interaction sessions. This complements work in robotics on developing safe robot motion and navigation planners by other partners within the COGNIRON project and elsewhere [Roy and Thrun, 2002]
4. The advantages of different types of video cameras are discussed, and we suggest that if using network based video cameras, it is wise to use at least one videotape-based camera as a backup in case of network problems, and vice versa. We also suggest some (obvious) ways to optimise camera placement and maximise coverage.
5. Similarly, we suggest it is good practice to have a backup robot available.
6. Sufficient time should be allocated to setup the experimental room and test all equipment and experimental procedures in situ. For example, our study used Radio Frequency (RF) based equipment to monitor and record the comfort level of the human subjects throughout the adult trial. We found that there was interference coming from sources that were only apparent when all the trial equipment was operating simultaneously.
7. Some points to consider when designing questionnaires are made. Completing questionnaires away from the trial area may conserve resources but influence the questionnaire results.
8. A careful consideration of methodological and design issues regarding the preparation of any user study will fundamentally impact any results and conclusions that might be gained.

It is vital that sufficient time is allowed for piloting and testing any planned trials properly in order to identify deficiencies and make improvements before the trials start properly. Full scale pilot studies will expose problems that are not apparent when running individual tests on the experimental equipment and methods. In our own studies the problems we did encounter were not serious enough to damage or invalidate major parts of the trials. We have highlighted other features of our trials we can improve upon, and made suggestions as to how to overcome the problems we have encountered. The lessons learned can be used to improve future trials involving human-robot interaction.

Acknowledgements

The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. Many thanks to Christina Kaouri, René te Boekhorst, Chrystopher Nehaniv, Iain Werry, David

Lee, Bob Guscott and Markus Finke for their help in designing, implementing and carrying out the trials.

References

ActivMedia Robotics (2005). ActivMedia Robotics Interface for Applications (ARIA). http://robots.activmedia.com/.

BBC Science news (2004). Interactive games run by the COGNIRON team from the University of Hertfordshire at the Science Museum, London.http://news.bbc.co.uk/2/hi/technology/396 269stm 2004.

Brooks, R.A. (1991). Intelligence without representation. *Artificial Intelligence* Vol 47, pp. 139–159.

COGNIRON (2005). Website available at www.cogniron.org. 2005

Crown copyright, (2003). An introduction to health and safety. HSE Books .http://www.hse.gov.uk/pubns/indg259.pdf. ISBN 0 7176 2685 7. 2003

Crown copyright, (1998). Data Protection Act 1998. http://www.hmso.gov.uk/acts/acts1998/19980029.htm#aofs. ISBN 0 10 542998 8. 1998.

P. Dario, E. Guglielmelli, C. Laschi (2001). Humanoids and Personal Robots: Design and Experiments. *Journal of Robotic Systems* 18 (12), pp. 673–690.

T. Fong,, I. Nourbakhsh, K. Dautenhahn (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, Vol. 42, pp. 143-166.

A. Green, H. Hűttenrauch, K Severinson Eklundh (2004). Applying the Wizard of Oz Framework to Cooperative Service discovery and Configuration. Proc. *IEEE Ro-man 2004,* 13th IEEE International Workshop on Robot and Human Interactive, Oka-yama, Japan, IEEE Press.

P. Hinds, T. Roberts, H. Jones (2004). Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task.. *Human-Computer Interaction*, Vol. 19, pp.151-18.

T. Kanda, T Hirano, D Eaton (2004). Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human-computer Interaction*, Vol. 19, pp. 61-84.

D. Maulsby, S. Greenberg, R Mander (1993). Prototyping an intelligent agent through Wizard of Oz. *Proc.ACM SIGCHI Conference on Human Factors in Computing Systems, Amsterdam*, The Netherlands, ACM Press, pp. 277-284.

B. Robins, K. Dautenhahn, J. Dubowski (2004). Investigating Autistic Children's Attitudes Towards Strangers with the Theatrical Robot - A New Experimental Paradigm in Human-Robot Interaction Studies . *Proc. IEEE Ro-man 2004,* 13th IEEE International Workshop on Robot and Human Interactive. Oka-yama Japan, IEEE Press, pp. 557-562

N. Roy, S. Thrun. Motion planning through policy search (2002). *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, EPFL, Switzerland, pp. 2419–2425.

UPR AS/A/2 (2004). University of Hertfordshire Ethics Policy. http://wisdom.herts.ac.uk/research/Indexes2/usefullinks2.htm. 2004.

VICTEC (2003). VICTEC Project website available at http://www.victec.org, 2003.