# Designing HRI Experiments with Humanoid Robots: A Multistep Approach

*Nadine M. Homburg, Student*
*Edith-Stein School, Darmstadt, Germany*
nstockhomburg@yahoo.de

## Abstract

Robots have penetrated many areas of daily life, including increased uses of humanoid robots in personal and organizational settings such as health care, eldercare, and service encounters with customers. Little research examines humanoid robots in these professional settings, even though the human–robot interaction (HRI) is particularly critical in such contexts. On the basis of a literature review and experience from several experimental studies, this article offers some guidance for designing HRI experiments with humanoid robots. In addition to detailing major challenges associated with designing HRI studies, this article suggests important next steps for experimental research with humanoid robots, as well as implications for further study.

## 1. Introduction

Humanoid robots increasingly appear in the daily contexts of people's lives [1], assisting human users in various professional settings such as retailing and hospitality [2 3]. Nestlé has inserted hundreds of humanoid service robots onto shop floors to sell Nescafé in Japan, for example [4 5]. The idea "that robots could become an integral part of groups and teams has developed from a promising vision into a reality" [6]. Accordingly, human–robot interactions (HRI) attract considerable attention in the robotic research community in general and among researchers dedicated to humanoid robots in particular [7 8]. Most of this research relies on proxy technologies, such as robotic heads [9 10], animals [11], or pictures on screens [12], instead of actual humanoid robots. Studies also tend to address simple interactions (e.g., human responses to a robotic smile or movement), rather than more complex interactions, likely because of a lack of clarity about how to integrate humanoid robots into complex experiments marked by rich HRIs.

Such relatively recent studies also highlight the growing need to investigate humanoid robots in personal and professional settings. Existing experimental studies rely on different approaches to their experimental designs, and no systematic method for integrating humanoid robots into experimental research has emerged.

This article therefore draws on a design science perspective "to extend the boundaries of human or organizational capabilities by creating new and innovative artifacts" [13, p. 5]. Specifically, I attempt to provide guidance for researchers who plan to conduct experiments involving interactions between humanoid robots and humans, as well as at other information systems researchers who are interested in what might be learned from such experimental studies. This guidance is important to synthesize an existing body of research and stimulate critical thinking [14]. Specifically, it seeks to provide guidance regarding how to include a humanoid robot in an experimental design.

Elaborating on this view could increase the impact of future experiments involving HRI and help researchers avoid the mistakes or gaps of prior literature. Thus, this article first provides a structured literature review and critical reflection of existing studies. Furthermore, particular examples are presented to illustrate successful uses of experiments to address important issues related to HRI.

Such experiments can be designed to be both effective and efficient. Efficiency pertains to the process of the experiment; effectiveness implies sufficient internal and external validity, whereas "An experiment that lacks internal validity fails by providing a misleading indication of the relation between the dependent and independent variable, while an experiment that lacks external validity produces results that are (or at least should be) divorced from the motivation of the study" [15]. To design and conduct effective and efficient experiments that clarify HRI,

URI: http://hdl.handle.net/10125/50448
ISBN: 978-0-9981331-1-9

this study proposes that five critical steps must be considered, as shown in Figure 1.

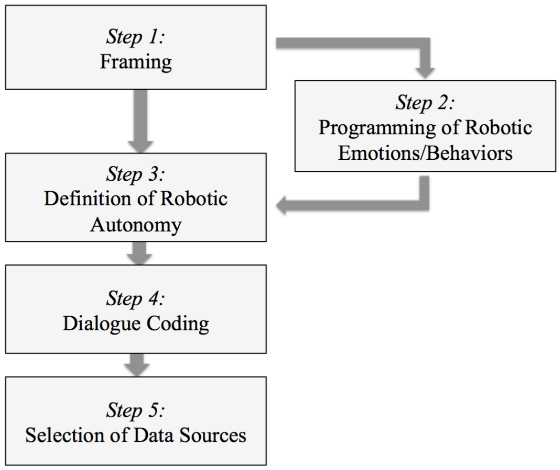



**Figure 1.** Multistep design model for experimental HRI research with humanoid robots

Accordingly, the next section provides an overview of extant research on experimental HRI with humanoid robots and proposes a structure for a multistep research design, as followed herein (Figure 1). Next, this article offers suggestions for dealing with the challenges that arise in various steps of experimental HRI research.

## 2. Experimental HRI Research with Humanoid Robots

Table 1 depicts selected studies on HRI according to their multistep research designs (Figure 1). The first step, *framing*, constitutes a critical choice that experimenters must make prior to starting the experiment. Rueben et al. [16, p. 435] claim that "the frame surrounding a given interaction could have comparable or even larger effects on judgments about that interaction than the independent variables typically studied in HRI research." Using framing criteria, prior research can be classified according to the scenario used (e.g., health care, household, hotel), information provided to participants (e.g., telling participants a fictional or the real purpose of the experiment), and the setting (laboratory, content-enriched laboratory, field). Most studies frame their experiments in specific, real-world scenarios, such as health care, elder care, home, classrooms, or hotels.

Prior research also features a variety of ways to present humanoid robots to participants. The possibilities vary from just showing a robot's head on a computer screen to having the participant interact with the actual humanoid robot in person.

Most studies are conducted in laboratory settings, which offer the advantages of minimizing the error factors that can arise in field studies. However, the external validity of these experiments is rather limited. As a compromise, researchers could rely on content-enriched settings, such as the environment of an apartment [17] or a hotel lobby [7].

Extant research also has *programmed robotic emotions and/or behaviors* in various ways. For example, some studies show participants pictures of faces on a screen [18 19]. Relying on the notion by Rueben et al. [16] that framing strongly influences the outcomes of an experiment, the robotic avatar presented in a given interaction likely exerts strong effects on the outcome of the HRI.

In terms of the *definition of robotic autonomy*, only one study has applied a completely autonomous mode in supporting the expressions of robotic emotions and/or behaviors [20]. Most studies adopt a semi-autonomous mode, in which the robot follows a pre-defined script and is operated by a human. Several studies apply a Wizard-of-Oz (WoZ) method, which "refers to a person (usually the experimenter, or confederate) remotely operating a robot, controlling any of a number of things, such as its movement, navigation, speech, gestures, etc." [21, p. 119]. Non-autonomous or semi-autonomous approaches, such as the WoZ method, have been criticized though (e.g., [22 23 24]). One methodological critique notes that the robot actually functions like a proxy in what are actually human–human interactions [25].

Most studies rely on predefined, standardized *dialogue* for their experiments. Despite the frequent use of dialogues, few studies explicitly report how they validated the dialogues that they present during the experiment ([7]).

Regarding the *data sources*, most studies rely on self-ratings [7 17 18 19 26 27 28 29 30 31 32 33 34]; see Table 1). Whereas self-ratings are useful to assess participants' characteristics, the assessment of emotions or behaviors based on self-ratings may be biased. For example, when participants respond to several questions about the manipulation and their emotional states, before and after the experiment, this method could invoke the threat of a single-source bias. This is because the same source provides assessments of both the independent variable (manipulation check) and the dependent variable.

Six studies contained in this literature review gather third-rater data. In these methods, the emotional and/or behavioral responses of participants are assessed by either an experimenter [31 32 35] or other independent raters [20 29 31 32 36]. Three studies use objective data (location tracking data [20], eye tracking data [26 30]) to assess human responses to robots. Both third-rater assessments and physiological data likely reduce common method bias [37 38].

| Author/s (year) | Framing (scenario/ depection/setting) | Robotic emotions / behaviors | Robotic autonomy | Dialogue | Data sources |
|---|---|---|---|---|---|
| Broadbent et al. (2007) | Healthcare/yes/laboratory | Face on screen (neutral) | Non-autonomous (pre-programmed) | - | Self-ratings |
| Eyssel et al. (2012) | -/yes/laboratory | - | Non-autonomous (pre-programmed) | Standardized phrase | Self-ratings |
| Heerink et al. (2010) | Elder care institution/no/field | Facial expressions, voice, gestures | Semi-autonomous (Wizard of Oz) | Standardized dialogue | Self-ratings, Independent rater |
| Kuo et al. (2009) | Healthcare/no/field | 3D face on screen | Non-autonomous (pre-programmed) | Standardized step by step instructions | Self-ratings |
| Mathur and Reichling (2016) | -/no/laboratory | Faces on screens | (not applicable) | - | Self-ratings, eye tracking |
| Miskam et al. (2012) | Healthcare/no/laboratory | Voice and gestures | Non-Autonomous (pre-programmed) | Standardized modules | Experimenter |
| Pitsch and Wrede (2014) | Education/ /field | Voice and gestures | Autonomous | Standardized modules | Independent rater |
| Rani et al. (2004) | Home /no/laboratory | Robotic movements | Semi-autonomous (robot has capability) | - | Self-ratings |
| Sakamoto et al. (2007) | -/no/laboratory | Movements, mimics, voice | Semi-autonomous (remote contr./operation) | Standardized dialogue | Self-ratings, eye tracking |
| Staffold et al. (2010) | Healthcare/no/field | - | Non-autonomous (pre-programmed) | Standardized modules | Self-rating |
| Stock and Merkle (2017) | Hotel/yes/laboratory (content-enriched) | Gestures, mimics, and speech | Semi-autonomous | Standardized dialogue | Self-ratings |
| Tanaka (2006, 2012) | Classroom/no/field | - | Semi-autonomous (upper body autonomous) | - | Experimenter, Independent rater |
| Tapus et al. (2012) | Healthcare/no/laboratory | Facial expressions and gestures | Non-autonomous | - | Independent rater |
| Woods et al. (2006) | Home/no/laboratory (content-enriched "Robot House") | Movements and gestures | Semi-autonomous (Wizard of Oz) | - | Self-ratings |
| Xu et al. (2013, 2014) | -/yes/laboratory | Positive and negative emotions | Non-autonomous | - | Self-ratings |

**Table 1.** Selected experimental HRI research with humanoid robots or robotic avatars

The next five sections detail different challenges, captured by the multistep approach in Figure 1, as well as ways they might be addressed. Furthermore, this article provides concrete examples of dealing with these challenges.

## 3. Framing

A common challenge in operationalizing independent variables is deciding how to frame the experimental study. Framing is particularly important for experiments that focus on how participants make decisions using cognitive processes and knowledge developed in response to their real-world education, training, and experience. Without relatively realistic stimuli, participants may not rely on such cognitive processes or leverage their knowledge of interest.

Two alternative, pertinent experimental settings might appropriately reflect professional situations in which humanoid robots and human users participate. The first is HRI during service encounters. As professional service providers, humanoid robots have manifold tasks, ranging from carrying customers' items and providing transportation services to welcoming and checking in consumers or answering routine questions. For example, the Japanese travel agency HIS runs the Henn-na Hotel almost completely with robots, which function as receptionists, luggage carriers, and room service personnel [39].
Another important setting is participation by humanoid robots in groups or teams. The wide availability of affordable humanoid robots has increased their use in small teamwork settings in industry [22], as well as larger group settings at conferences [23]. More sophisticated robots also support complex teamwork projects across a wide range of settings, such as search-and-rescue missions [24] or space exploration [25]. In a recent study, Gombolay and Sturla [5] compare the effectiveness of groups directed by humanoid robots against those directed by human supervisors. Surprisingly, participants in the robotic group reported higher satisfaction and achieved higher effectiveness scores than participants in the human-led group.

In general, researchers should try to make the setting as realistic as possible, which in many cases means using an actual humanoid robot. The presentation of the robot must reflect the research goals and characteristics of the participants. Figure 2 depicts the two alternative sample settings, with a humanoid robot (Pepper) included in the experiment as a customer service provider (A) or group work supervisor (B).

(A) Robot as receptionist [6]

(B) Robot as supervisor

**Figure 2.** Experimental settings with the humanoid robot Pepper in different professional roles

In an interaction such as the one depicted in Figure 2A, the setting should be possible and realistic with regard to using a robot. For example, it might seem unrealistic to use a robot instead of a self-service technology at a check-in counter, because this service requires little interaction. Furthermore, humanoid robots generally function to replace humans, not other service technologies. Once the setting is chosen, the surroundings need to be as authentic as possible but also adjusted to the presence of the robot. In Figure 2A, the robot stands next to, rather than behind, the reception desk, which enables the customer to see what the robot is displaying on its tablet or sign the tablet. In addition, the participant can clearly recognize the robot's gestures and postures, which is important for the effectiveness of the manipulations (i.e., emotions).

Figure 2B depicts a robot in a group setting, positioned in front of a group to provide directions to the group members and support the fulfillment of their task. In this experiment, the participants had to design a low-energy building and build a prototype with Lego®-bricks. The robot observed their group activities and provided regular feedback.

# 4. Programming Robotic Emotions/Behaviors

## 4.1 Conceptual Basis for Emotion and Behavioral Programming

Robotic emotions and behaviors have become an increasingly important element of experimental HRI research but also one of the most critical challenges for these experiments. A fruitful way to identify authentic robotic emotions or behaviors is to adopt psychological theories as a conceptual basis. Psychological research identifies a range of emotions—anger, disgust, fear, joy, sorrow, and surprise—that define human–human interactions [40]. Subsequent research has proposed a circumplex model of emotions that encompasses two orthogonal dimensions [41 42]: the hedonic valence of pleasant versus unpleasant and arousal (i.e., low vs. high activation), as depicted in Figure 3.

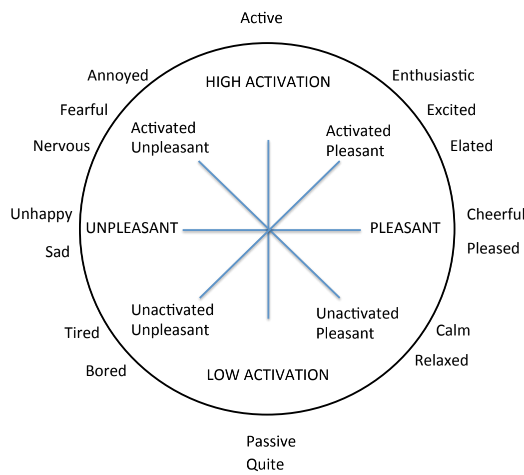



**Figure 3.** Circumplex model of emotions as a conceptual basis for programming robotic emotions [41]

The circumplex model of emotions and behaviors has been well established in psychological research and applied in previous robotic research, such as when Breazeal [43] structured the basic emotional and behavioral expressions of the humanoid robot Kismet according to this model (Figure 4).

During human–human interactions, emotions can be expressed by voice, face, gesture, and posture [44]. That is, emotions and behaviors typically are communicated among humans through vocal, facial, and bodily expressions [45]. In contrast, humanoid robots often are limited in the ways they can signal emotions. For example, neither the Pepper nor the NAO models can express facial expressions, though some robots, such as the Kismet [40], express manifold facial expressions. Therefore, depending on which robot is used, different possibilities exist for programing emotions and behaviors to varying extents.

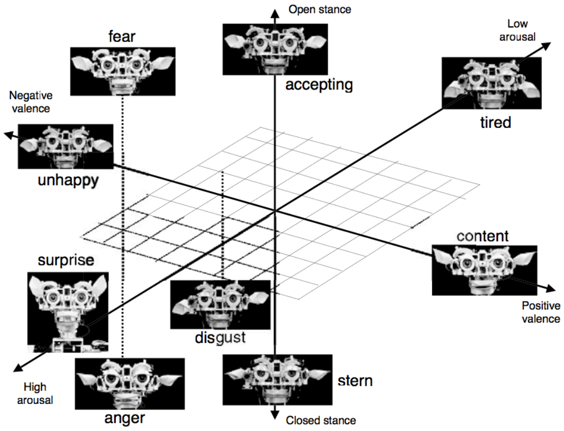



**Figure 4**. Emotions of the Kismet robot according to the circumplex model of emotions [43]

In terms of the robot's *verbal expressions*, researchers might cause it to express emotions through certain voice pitches and words being said. In this case, it is important that the robot sounds natural, which can be challenging for tones such as laughter. In addition, a setting can quickly come to seem unnatural if the robot uses too many words. The choice of words, including their order and the speed of the robot's verbal expressions, thus must be chosen carefully to ensure an effective manipulation.

Sometimes researchers have the additional possibility of signifying emotions with the robot's *facial expressions*. Many humanoid robots, such as NAO or Pepper, only offer simple, moderate facial features; their LED head features a graphical face. Experimenters generally can vary the robot's eye color or head movements. The challenge is to program clearly recognizable emotions and behaviors. However, overemphasizing certain expressions can quickly become unnatural and even impede HRI or frighten participants.

If researchers use a humanoid robot with a full body, *bodily expressions* offer a rich source for emotion and behavioral expressions. Here, researchers can rely on knowledge about bodily expressions by humans that are easy for other human participants to recognize. An important aspect relates to the attitudes of participants toward robots. Extant research indicates that humans have robot anxiety; they associate negative characteristics and behaviors or even threats with robots. Extensive movements by a humanoid robot during the experiment might foster such anxiety and perhaps impede the effectiveness of the experiment.

### 4.2 Emotion Validation

Regardless of how researchers choose to design the robot to express emotions, a validation test is necessary. In particular, the robot's non-verbal and verbal expressions should be assessed by independent raters, such as a set of potential participants, who indicate which emotion they perceive the robot to have shown. The manipulation of these emotions should be adjusted until the validation test achieves an accuracy rate of greater than 80%.

Stock and Merkle [7] use a stepwise approach to identify appropriate positions for NAO to express emotions through its output behavior [7, 8]. First, the design reflected extant literature in psychology [46] and robotic research that suggests various behavioral outputs of emotional expressions [47]. Second, a web search sought to identify 100 pictures for each of the five emotions. Using these pictures, the two most typical bodily expressions for happiness, surprise, anger, and frustration, as well as one neutral position, were programmed. Third, the programmed positions were presented to 234 students (18–43 years of age; 67% men; 80% technical background), who had to rate the bodily expressions exhibited by NAO. They readily identified the bodily expressions for all four valenced emotions, whether pleasant (happiness 91%; positive surprise 95%) or unpleasant (anger 83%; frustration 94%). In addition, the neutral emotion expression by the robot was recognized by 93% of these students. In the subsequent main experiment, each of the five emotions expressed by NAO's body gestures was displayed during the HRI [29], as depicted in Figure 5.

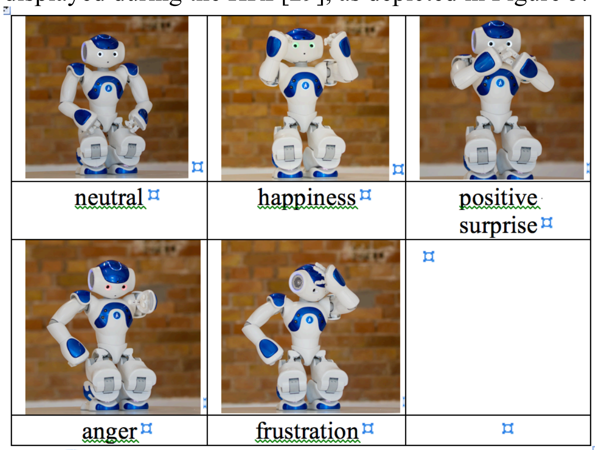



**Figure 5**. Emotion expressions by NAO robot during a human–robot interaction experiment

## 5. Definition of Robotic Autonomy

To define robotic autonomy, researchers face several challenges, and no resolution has been achieved. Most humanoid robots lack a well-developed voice recognition system, such that the robot often cannot understand human participants' verbal expressions. This issue might be especially challenging for experiments performed in loud settings, such as group or field tests. Robotic face recognition systems also are far from allowing robots to recognize or respond to humans' facial expressions.

Even if a robot has a sophisticated voice recognition system and can understand a human participant's verbal expressions, a challenge remains with regard to prompting the robot to respond in a similar way during the experiment. This goal is important though, to ensure that human responses to the HRI are comparable.

In terms of robot autonomy, three designs emerge from extant literature: The robot acts freely without any influence of an operator during the experiment (autonomous), robotic movements and/or phrasings are completely pre-programmed (non-autonomous), or the robot acts according to a pre-programmed structure but is partly controlled by an operator (semi-autonomous).

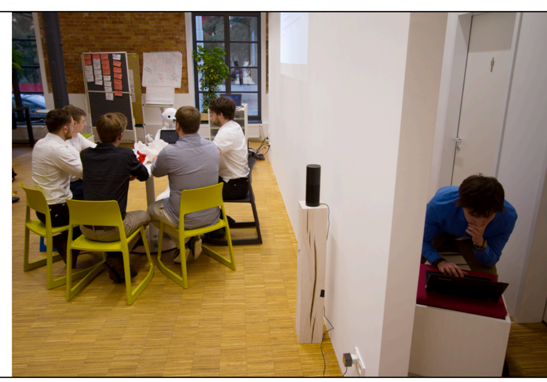
When using the autonomous mode, researchers must realize that participants have a variety of ways to express their needs or articulate solutions to the task at hand. Therefore, the robot needs access to a rich, complex set of potential answers so that it can engage in natural conversation during the HRI. Cloud solutions, such as IBM's Watson, have not achieved sufficient voice recognition speed or related robotic responses to ensure a natural conversion. A programmer thus would have to program all possible movements and/or phrasings into a pre-programmed HRI—a task that quickly grows more difficult as the experiment grows longer and more complicated.

When using the non-autonomous or semi-autonomous mode, extant research frequently relies on the WoZ method ([10]). To apply this method, researchers tell participants that the humanoid robot acts autonomously, but in reality, the robot is (partly or completely) operated by a human, hidden behind a curtain or wall. Figure 6 depicts some sample applications for the WoZ method. A Wizard can be employed to guarantee some standardized repertoire of movements, gestures, or phrases, given by the robot. This standardization is important for enabling comparisons of HRI across a set of experiments. The Wizard also can make the robot respond seemingly spontaneously to unexpected movements or phrases issued by the human participants. In this case, a human–human interaction is taking place through a robot. But researchers need to take care to avoid this scenario of a human–human conversation occurring through a robot. Furthermore, ethical concerns arise, in terms of social deception ([23]). Extant literature suggests several criteria that should be reported explicitly in research papers, to avoid the disadvantages of the WoZ method ([23 48]). If these criteria are met, the WoZ method can offer important advantages, in terms of experimental effectiveness.

First, robotic behaviors can be standardized across experiments, which helps ensure the internal and external validity of the findings. Otherwise, the experimental procedure could be interrupted by random mistakes by the robot. Second, the robot can respond quickly to verbal or non-verbal expressions by the participant that might not have been considered prior to the experiment. In turn, the natural flow of conversation during the experiment improves.

With a semi-autonomous method, researchers would combine the autonomous mode and the WoZ method in an experiment, to ensure validity and dialogue that barely differs across each experiment. In this case, it is important to pre-program the answers that the Wizard may use in the conversation. If necessary, the Wizard can add small comments, in response to the participant, but the central conversation and word use by the robot will stay the same.

(A) Robot as receptionist



(B) Robot as supervisor

**Figure 6.** The Wizard of Oz method in various experimental settings

## 6. Dialogue Coding

For dialogue coding, coders must deal with several, previously unpredictable challenges to create a natural flow in the HRI, especially if the robot is speaking autonomously. After the dialogue coding, researchers should run preliminary tests and improve the coding and dialogue for the main experiment. During the process of coding the dialogue, several key aspects need to be considered, as detailed next.

### 6.1 Varying Answers

The answers and words expressed by the robot must vary for the interaction to feel natural, especially for simple phrases such as "yes," "no," "you're welcome," or "thanks"—that is, words that are repeated several times in conversation throughout the experiment. Instead of using one version to say the same thing over and over again, researchers should invoke variations. Otherwise, the conversation starts to sound unnatural, which impedes experimental effectiveness. In addition, variation allows human participants to respond in a more natural manner.

### 6.2 Starting the Interaction

When entering the experimental setting, participants might focus strongly on the robot or expect the robot to start the interaction. The robot then needs to be able to start the interaction without requiring a long pause, which would create an awkward situation. This goal is easier to achieve with a semi-autonomous or non-autonomous method, but even then, experimenters and Wizards must act quickly. A good way to start the interaction is to have the robot ask the participant a couple of (warm-up) questions, such that the participant becomes familiar with the situation and the robot. This introduction is important; for most participants, it is their first time interacting with a robot. Encouraging the participant to answer a couple of questions often sparks a natural conversation.

### 6.3 Repetition and Stops

Regardless of the degree of autonomy, the robot should always be able to repeat everything it says and, in some cases, rephrase things. This ability is essential for a natural conversation and to prevent awkward pauses. Not being able to understand everything the robot says can lead participants to feel uncomfortable or even frighten them. In addition, the robot must be able to stop and explain things better, such as when asked a question by the participant. This element of a natural interaction also helps prevent an awkward situation in which the participant asks a question and the robot fails to acknowledge it, and instead just keeps going with what it was saying before.

### 6.4 Prosody

The need to understand the robot well leads to the next challenge, namely, getting the robot to speak at the right pace and pitch, with pauses in the right places. Experimenters should listen to the programmed dialogue over and over, until it sounds natural and the robot can be understood easily. If the robot strings together several sentences at once, coders should add pauses after every two or three sentences. A human might not include these pauses while talking, but they help participants in HRI. The length of these pauses should be long enough for the participant to process the information but not so long that the participant starts talking in between. Poor adjustments of these pauses could lead participants to feel uncomfortable, which could impede the effectiveness of the experiment.

## 7. Selection of Data Sources

Participants can provide various measures. This section differentiates among self-rating, third-rater assessments, and physiological measures.

### 7.1 Self-Ratings

To improve the validity of self-ratings, researchers need to ask participants to answer spontaneously, to avoid socially desired answers. In addition, researchers should use established psychological measures for the participants.

In some cases, such as for participants diagnosed with autism or small children, self-ratings are difficult or impossible to use. Thus, experimental research increasingly relies on third-rater or physiological data.

### 7.2 Third-Rater Assessments

When employing independent raters, Bartel [41] suggests using at least three independent raters, who can assess emotional and behavioral responses in experimental settings by watching video recordings. Extant research recommends a two-way rating approach that requires each rater to provide a score for each measure [37 38]. Raters should initially jointly reach consensus about their assessments. Then they can perform independent ratings of participants' behaviors during the experiment, with no consultation about the ratings throughout the rating process [49].

### 7.3 Physiological Measures

Issues with other forms of data create a need for additional measurements to validate ratings by participants or experimenters. Human responses to HRIs can be measured with psychological or physiological indicators. To take such measures, researchers have several options, depending on what they want to measure and at which level of detail. Laboratory studies often rely on questionnaires to measure participants' psychological reactions, such as the Positive and Negative Affect Scale [30] or stationary devices that measure participants' physiological responses. Physiological responses, and in particular heart rate (HR) and heart rate variability (HRV), are good indicators of participants' arousal [50].

An important decision relates to the use of stationary versus mobile physiological measurement devices. Despite the advantages of stationary physiological measurements, such as good accuracy and software with comprehensive functionality [51], their application is typically restricted to controlled laboratory settings in which the participants cannot move freely. In contrast, wearable devices allow for mobile physiological measurements. These devices are small and need little space [52]; they mostly rely on photoplethysmography sensors to measure cardiovascular indicators [52]. For experiments with HRIs, wearables are promising; they are easy to handle and do not restrict the movements of the participants. In turn, for data analysis, it is important to compare the measures during the experiment with measures taken before the experiment (baseline measure).

## 8. Discussion

Following these steps to design experimental research with robots can increase the internal and external validity of research in several respects.

First, realistic stimuli and experimental designs motivate participants to behave in authentic ways and avoid socially desired behaviors. Second, clearly programmed and empirically tested robotic emotions and/or behaviors help avoid confounding effects. In particular, confounding effects due to a lack of differentiation between the control group and experimental group can be avoided. The extent to which this step needs to be applied depends on the experiment. Third, a clearly defined dialogue limits experimenter bias, particularly when the behaviors of the humanoid robot are defined by a concrete script of verbal expressions and standardized with the WoZ method. Fourth, data from multiple sources reduce the risk of a single-source bias.

This research also provides insights for testing hypotheses about HRI in professional settings. Further research should apply this multistep model to develop experiments with humanoid robots in various professional roles, such as team member, subordinate, or even supervisor in organizations, as well as frontline personnel at the boundary with customers.

An important action, prior to starting the main experiment, is a *pilot test*. With a set of pilot experiments, researchers can increase internal and external validity through the sufficient execution of the multistep approach (Figure 1). For example, they can test whether the experimental framing is adequate, realistic, and logically consistent and if the chosen autonomy mode supports the experimental procedure sufficiently.

This article focuses on humanoid robots in professional settings, but moving forward, research could apply this multistep model to investigate personal robots. In contrast with professional settings, personal robots serve humans in private spheres.

In addition, I attempt to provide guidance for experimental research with humanoid robots. The suggestions should be considered with care for other research designs; their relevance may be restricted for non-humanoid robots, such as industrial forms.

By presenting this multistep approach for experimental research with humanoid robots, this article seeks to increase the impact of future HRI experiments and help other researchers avoid the mistakes of prior literature.

## 9. References


[1] N. Reich, and F. Eyssel. Attitudes Towards Service Robots in Domestic Environments: The Role of Personality Characteristics, Individual Interests, and Demographic Variables, Journal of Behavioral Robotics, vol. 4:2, pp. 123-130, 2013.

[2] M.-J. Han, C.-H. Lin, and K.-T. Song. Emotional Expression Generation Based on Mood Transition and Personality Model, IEEE Transactions on Cybernetics, vol. 43:4, pp. 1290-130, 20133.

[]3 T. Zhang, B. Zhu, L. Lee, and D. Kaber. Service Robot Anthropomorphism and Interface Design for Emotion in Human-Robot Interaction, 4th IEEE Conference on Automation Science and Engineering, Washington, DC, 2008.

[4] M. F. Jung, F. Eyssel, S. Šabanović, S., and M. Fraune. Robots in Groups and Teams, CSCW '17 Companion, February 25 - March 01, 2017, Portland, OR, 2008. ACM 978-1-4503-4688-7/17/02.

[5] Nestlé. Nestlé to Use Humanoid Robot to Sell Nescafé in Japan, 2016, [http://www.nestle.com/ media/news/nestle-humanoid-robot-nescafe-japan].

[6] G. Gombolay and S. Sturla. Decision-Making Authority, Team Efficiency and Human Worker Satisfaction in Mixed Human-Robot Teams. In: Autonomous Robots, 39, 293-312, 2015.

[7] R. Stock and M. Merkle. A Service Robot Acceptance Model: User Acceptance of Humanoid Robots During Service Encounters, PERCOM Conference, Hawaii, 2017.

[8] R. Stock. Emotion Transfer from Frontline Social Robots to Human Customers During Service Encounters: Testing an Artificial Emotional Contagion Model, International Conference on Information Systems, Dublin, Ireland, 2016.

[9] C. Breazeal, A. Wang, and R. Piard. Experiments with a Robotic Computer: Body, Affect and Cognition Interactions, HRI 07, Arlington, Virginia, USA, 2007.

[10] L. D. Riek, P. C. Paul, and P. Robinson. When My Robot Smiles at Me: Enabling Human-Robot Rapport via Real-time Head Gesture Mimicry, Journal on Multimodal User Interfaces, vol. 3:1-2, pp. 99-108, 2010.

[11] M. P. Michalowski, S. Sabanovic, and H. Kozima. A Dancing Robot for Rhythmic Social Interaction, HRI 07, Arlington, Virginia, USA, 2007.

[12] M. Hashimoto, M. Yamano, and T. Usu. Effects of Emotional Synchronization in Human-Robot KANSEI Communication, 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 2009.

[13] A. R. Hevner, S. T. March, J. Park, J., and S. Ram, Design Science in Information Systems Research, vol. 28:1, pp. 75-105, 2004.

[14] I. Benbasat and R. W. Zmud. Empirical Research in Information Systems: The Practice of Relevance, MIS Quarterly, vol. 23:1, pp. 3-16, 1999.

[15] R. Libby, R. Bloomfield, and M. W. Nelson. Experimental Research in Financial Accounting, Accounting Organizations and Society, vol. 27, pp. 794, 2002.

[16] M. Rueben, F. J. Bernieri, C. M. Grimm, and W. D. Smart. Framing Effects on Privacy Concerns About a Home Telepresent Robot. HRI, March 06-09, Vienna/Australia, 2017.

[17] S. N. Woods, M. L. Walters, K. L. Koay, and K. Dautenhahn. Methodological Issues in HRI: A Comparison of Live and Video Based Methods in Robot to Human Approach Direction Trials, 15th IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, UK, September 6-8, 2006.

[18] E. Broadbent, B. MacDonald, L. Jago, M. Juergens, and O. Mazharullah. Human Reactions to Good and Bad Robots. In Intelligent Robots and Systems, 2007. IEEE/RSJ International Conference on (pp. 3703-3708). IEEE.

[19] I. H. Kuo, J. M. Rabindran, E. Broadbent, Y. I. Lee, N. Kerse, N., R. M. Stafford, and B. A. MacDonald. Age and Gender Factors in User Acceptance of Healthcare Robots. In Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on (pp. 214-219). IEEE.

[20] K. Pitsch and S. Wrede. When a Robot Orients Visitors to an Exhibit. Referential Practices and Interactional Dynamics in Real World HRI. In Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on (pp. 36-42). IEEE.

[21 L. D. Riek. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines, Journal of Human-Robot Interaction, vol. 1:1, pp. 119-136 2012.

[22] C. Breazeal. Social Interactions in HRI: The Robot View. IEEE Transactions on Systems, Man, and Cybernetics vol. 34, pp. 181-186, 2004.

[23] N. M. Fraser and G. N. Gilbert. Simulating Speech Systems. Computer Speech & Language, vol. 5:1, pp. 81–99, 1991.

[24] L. D. Riek and R. Watson. The Age of Avatar Realism. IEEE Robotics & Automation Magazine, vol. 17:4, pp. 37–42, 2010.

[25] A. Weiss. Validation of an Evaluation Framework for Human-Robot Interaction. The Impact of Usability, Social Acceptance, User Experience, and Societal Impact on Collaboration with Humanoid Robots. Unpublished doctoral dissertation, University of Salzburg, 2010.

[26] M. B. Mathur and D. B. Reichling. Navigating a Social World with Robot Partners: A Quantitative Cartography of the Uncanny Valley. Cognition, vol. 146, pp. 22-32, 2016.

[27] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel. 'If You Sound Like Me, You Must be More Human': On the Interplay of Robot and User Features on Human-Robot Acceptance and Anthropomorphism. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction 2012, (pp. 125-126).

[28] P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby. Anxiety detecting robotic system–towards implicit human-robot collaboration. Robotica, vol. 22:1, pp. 85-95, 2004.

[29] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. Relating Conversational Expressiveness to Social Presence and Acceptance of an Assistive Social Robot. Virtual Reality, vol. 14:1, pp. 77-84, 2010.

[30] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita. Android as a Telecommunication Medium with a Human-Like Presence. Human-Robot Interaction, 2007.

[31] F. Tanaka and S. Matsuzoe. Children Teach a Care-Receiving Robot to Promote their Learning: Field Experiments in a Classroom for Vocabulary Learning. Journal of Human-Robot Interaction, vol. 1:1, 2012.

[32] F. Tanaka, J. R. Movellan, B. Fortenberry, and K. Aisaka. Daily HRI Evaluation at a Classroom Environment: Reports from Dance Interaction Experiments. In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction (pp. 3-9). ACM, 2006.

[33] J. Xu, J. Broekens, K. Hindriks, and M. A. Neerincx. Bodily Mood Expression: Recognize Moods from Functional Behaviors of Humanoid Robots. In International conference on social robotics (pp. 511-520). Springer, Cham, 2013.

[34] J. Xu, J. Broekens, K. Hindriks, and M. A. Neerincx. Effects of Bodily Mood Expression of a Robotic Teacher on Students. In Intelligent Robots and Systems 2014 IEEE/RSJ International Conference on (pp. 2614-2620). IEEE.

[35] Miskam, M. A., M. A. Hamid, H. Yussof, S. Shamsuddin, N. A. Malik, and S. N. Basir. Study on Social Interaction Between Children with Autism and Humanoid Robot NAO, Applied Mechanics and Materials, vol. 393, pp. 573-578, 2013.

[36] A. Tapus, A. Peca, A. Aly, C. Pop, L. Jisa, S. Pintea, ... and D. O. David. Children with Autism Social Engagement in Interaction with Nao, an Imitative Robot: A Series of Single Case Experiments. Interaction Studies, vol. 13:3, pp. 315-347, 2012.

[37] J. Baer, J. C. Kaufman, and C. A. Gentile. Extension of the Consensual Assessment Technique to Nonparallel Creative Products, Creativity Research Journal, vol. 16:1, pp. 113-117, 2004.

[38] S. G. Barsade. The Ripple Effect: Emotional Contagion and its Influence on Group Behavior. Administrative Science Quarterly, vol. 47:4, pp. 644-675, 2002.

[39] M. Rajesh. Inside Japan's First Robot-Staffed Hotel, Japan Holidays [http://www.theguardian.com/travel/2015/aug/14/japan-henn-na-hotel-staffed-by-robots], 2015.

[40] P. Ekman and H. Oster. Review of Research 1970 to 1980. In: Ekman, P. (Ed.), Emotion in the Human Face. Cambridge University Press, Cambridge, UK, pp. 147–174, 1982.

[41] C. Bartel and R. Saavedra. The Collective Construction of Work Group Moods, Administrative Science Quarterly, vol. 45:2, pp. 197-231, 2000.

[42] R. J. Larsen and E. E. Diener. Promises and Problems with the Circumplex Model of Emotion. In M. S. Clark (ed.), Review of Personality and Social Psychology: Emotion and Social Behavior, vol. 114, pp. 25-59. Newbury Park, CA: Sage, 1992.

[43] C. Breazeal. Emotion and Sociable Humanoid Robots. International Journal of Human-Computer Studies, vol. 59, pp. 119-155, 2003.

[44] R. Levenson. Human Emotions: A Functional View. In: Ekman, P., Davidson, R. (Eds.), The Nature of Emotion. Oxford University Press, New York, NY, pp. 123–126, 1994.

[45] T. Bänzinger, D. Grandjean, and K. Scherer. Emotion Recognition from Expressions in Face, Voice, and Body: The Multimodal Emotion Recognition Test (MERT), Emotion vol. 9:5, pp. 691-704, 2009.

[46] A. Pease. Body Language: How to Read Others' Thoughts by their Gestures, Great Britain: University Printing House, Oxford, 1981.

[47] S. Shamsuddin, H. Yussof, L. I. Ismail, S. Mohamed, F. A. Hanapiah, and N. I. Zahari. Initial Response in HRI - A Case Study on Evaluation of Child with Autism Spectrum Disorders Interacting with a Humanoid Robot Nao, Procedia Engineering, vol. 41, pp. 1448-1455, 2012.

[48 A. Green, K. S. Eklundh, and B. Wrede. Integrating Miscommunication Analysis in Natural Language Interface Design for a Service Robot. In IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2006.

[49] T. M. Amabile. The Social Psychology of Creativity: A Componential Conceptualization, Journal of Personality and Social Psychology, vol. 45:2, pp. 357-376, 1983.

[50] G. G. Berntson, J. T. Bigger, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagarja, S. W. Porges, J. P. Saul, P. H. Stone, and M. W. van der Molen. Heart Rate Variability: Origins, Methods, and Interpretive Caveats, Psychophysiology, vol. 34:6, pp. 623-648, 1997.

[51] N. Sharma, and T. Gedeon. Objective Measures, Sensors and Computational Techniques for Stress Recognition and Classification: A Survey, Computer Methods and Programs in Biomedicine, vol. 108:3, pp. 1287–1301, 2012.

[52] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida. Wearable Photoplethysmographic Sensors—Past and Present, Electronics, vol. 3, pp. 282–302, 2014.