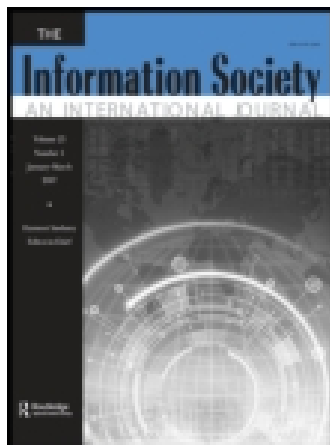


On: 02 June 2015, At: 03:35

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Information Society: An International Journal

Publication details, including instructions for authors and subscription information:
<http://www.tandfonline.com/loi/utis20>

Integrating Constrained Experiments in Long-Term Human-Robot Interaction Using Task- and Scenario-Based Prototyping

Dag Sverre Syrdal^a, Kerstin Dautenhahn^a, Kheng Lee Koay^a & Wan Ching Ho^a

^a Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, United Kingdom

Published online: 13 May 2015.



[Click for updates](#)

To cite this article: Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay & Wan Ching Ho (2015) Integrating Constrained Experiments in Long-Term Human-Robot Interaction Using Task- and Scenario-Based Prototyping, *The Information Society: An International Journal*, 31:3, 265-283, DOI: [10.1080/01972243.2015.1020212](https://doi.org/10.1080/01972243.2015.1020212)

To link to this article: <http://dx.doi.org/10.1080/01972243.2015.1020212>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Versions of published Taylor & Francis and Routledge Open articles and Taylor & Francis and Routledge Open Select articles posted to institutional or subject repositories or any other third-party website are without warranty from Taylor & Francis of any kind, either expressed or implied, including, but not limited to, warranties of merchantability, fitness for a particular purpose, or non-infringement. Any opinions and views expressed in this article are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor & Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

It is essential that you check the license status of any given Open and Open Select article to confirm conditions of access and use.

Integrating Constrained Experiments in Long-Term Human–Robot Interaction Using Task- and Scenario-Based Prototyping

Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Wan Ching Ho

Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, United Kingdom

In order to investigate how the use of robots may impact everyday tasks, twelve participants in our study interacted with a University of Hertfordshire Sunflower robot over a period of 8 weeks in the university's Robot House. Participants performed two constrained tasks, one physical and one cognitive, four times over this period. Participant responses were recorded using a variety of measures including the System Usability Scale and the NASA Task Load Index. The use of the robot had an impact on the experienced workload of the participants differently for the two tasks, and this effect changed over time. In the physical task, there was evidence of adaptation to the robot's behavior. For the cognitive task, the use of the robot was experienced as more frustrating in the later weeks.

Keywords assistive robotics, domestic robots, human–robot interaction, prototyping

In the field of human–robot interaction, domestic, human-centered environments present serious challenges for prototyping human–machine interactions. In particular, when addressing future and emergent technologies, it is a challenge to enable interactions that are situated in such a way that they are meaningful to the user, and allow users to translate this experience to their everyday life. Moreover, the experience of such interactions is subjective, and the relationship between interactants,

technologies, and situations can be complex and dynamic (Buchenau and Suri 2000). On the technical side, cutting-edge technologies often do not have the stability required to function autonomously in an effective and safe manner for sustained periods of time outside of highly constrained settings. However, such feedback is critical for guiding the development of these technologies. This necessitates a high degree of pragmatism and creativity when developing appropriate methodologies for examining how prospective users interact with these technologies, and how these interactions may benefit or hinder the user (Dautenhahn 2007).

While there have been studies of actual robots acting autonomously in a domestic environment without continuous oversight by experimenters, either the robots employed have had limited movement capabilities, and served mainly as physically embodied conversational agents (not unlike those described in Bickmore and Cassell 2005) as in the KSERA project (Payr 2010), or the robots were market-ready products (Fernaes et al. 2010; Sung et al. 2008) or at a late stage in the development cycle (Kidd and Breazeal 2008). Furthermore, due to the cost in time and resources to set up and run the experiments, live interactions with robotic technologies in complex usage scenarios usually involve only a relatively small number of participants (Walters et al. 2011; Huijnen et al. 2011). While it is often desirable to run studies with the largest number of participants possible for greater generalizability, there is also the need for studies that allow for a wide range of interactions to capture data on human–robot interaction in all its richness. This balance lies at the heart of our efforts to develop, adapt, and use prototyping methodologies for domestic human–robot interaction (Syrdal et al. 2008).

PROTOTYPING OF HUMAN–ROBOT INTERACTION

Broadly, there are two different approaches to prototyping of human–robot interaction. The first one is a holistic,

© Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Wan Ching Ho.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named authors have been asserted.

Received 28 December 2013; accepted 15 August 2014.

Address correspondence to Dag Sverre Syrdal, Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom. E-mail: d.s.syrdal@herts.ac.uk

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/utis.

scenario-based approach (Carroll 2000), which takes a high-level view of the situations and tries to capture the experience of the interaction through narratives. Here the participants' interactions with the robot are framed within a narrative that allows them to evaluate the potential impact of the prototype in everyday life situations. These scenarios can be presented to the participants as written stories (Blythe and Wright 2006), videos (Walters et al. 2011; Syrdal et al. 2010), theater performances (Syrdal et al. 2011; Chatley et al. 2010; Newell et al. 2006), or live human-robot interactions (Koay et al. 2009). The second approach is more reductionist and condenses and abstracts the salient features of the interaction into a controlled experimental setup. This approach has been used successfully for studying human-robot proxemics (Tapus et al. 2008; Koay et al. 2007; Dautenhahn et al. 2006), specific robot behavior styles (Syrdal et al. 2009; Fussell et al. 2008; Bartneck et al. 2005), and different user groups.

These two approaches are not mutually exclusive. For instance, Walters et al. (2011) combined a high-level narrative with a highly constrained experimental manipulation in a video study. However, each has clear strengths and weaknesses when compared to the other. The narrative approach provides insights into how robotic technologies may impact on people's lives on a more conceptual level. It does not, however give the participant the clear ability to experience and differentiate between the ways that the particularities of a robot's behavior or characteristics impact specific interactions. Highly controlled, experimental studies, on the other hand, are often lacking in ecological validity, but allow for in-depth understanding of specific aspects of the interaction.

The study presented here fruitfully brought together both approaches: The controlled experiments were integrated with open-ended scenarios as part of a long-term study (Syrdal et al. 2014). These studies were conducted in the University of Hertfordshire Robot House.

UNIVERSITY OF HERTFORDSHIRE ROBOT HOUSE

The UH Robot House is a residential house, near the University of Hertfordshire campus, that has been adapted for human-robot interaction studies. It has been augmented into a "smart home" with low-cost, resource-efficient sensor systems that inform the robots about user activities and other events in the environment (Duque et al. 2013). Moreover, it offers ecological validity because it is a real working house, with kitchen appliances, a TV, a doorbell, and so on. Throughout the studies presented here, participants primarily used the living room, dining area, and kitchen, sometimes responding to events (visitors, deliveries, etc.) at the front door, with an extra room used in the briefing for the open-ended scenario. In general, the Robot House serves as an effective test bed for prototyping

domestic human-robot interactions. Its infrastructure supports interactions with a range of robots such as the UH Sunflower robot (Koay et al. 2013), PeopleBots (Walters et al. 2011), and the IPA Care-O-Bot 3 (Parlitz et al. 2008; Koay et al. 2014).

CONSTRUCTED PERSONAS

Personas are understood in human-computer interaction as fictional yet highly realized users of a given technology (Chang et al. 2008). By creating and extrapolating behaviors, goals, histories, and characteristics of these, it is possible to tightly focus the technological development. The specific personas used to guide the scenario development in the Robot House were a couple in their mid-to-late sixties. The personas were given work, interests, and health issues, which are summarized next.

The Husband (David) is recently retired from a white-collar profession. He is looking forward to spend some time focusing on his hobbies, which include reading, watching documentaries, and building military models. He has a heart condition, which requires him to take medication regularly. For some reason, he often forgets to take this medication and has to be reminded by his wife daily. He also has a condition (likely arthritis in the knees) that gives him some mobility issues.

The Wife (Judy) works from home most days. Her husband's recent retirement and associated distractions are causing her some stress, and the couple some tension. She normally stays in her home office almost exclusively during her working hours, interacting with David primarily at mealtimes. She is used to computing technology, relying on it to work effectively from her home office. This has also enabled her and David to maintain close contact (using Skype and other social media) with their children and grandchildren.

Based on the lives of these personas, we created a "typical" day comprised of episodes in which the robot was utilized to aid "Judy" and "David" in their daily activities. See Figures 1 and 2 for episodes from a scenario based on a "typical" day for the user personas. The evaluation scenarios were created by examining the possible roles that the robot could play in the different episodes that comprised a "typical" day for the two user personas. This was done both as high-level narrative-based interactions, which presented scenarios where the interaction with the robot was within a specific context for the participants, and through constrained and experimental examinations of the role of the robot within specific tasks.

OPEN-ENDED SCENARIOS

The open-ended scenarios sought to convey the impact of the agent within a wider context to the participants in an



FIG. 1. Episode from a “Normal Day” for the user personas (1).

evaluation study. To achieve this, two open-ended scenarios were created. They were inspired by the Persona Scenarios (as shown in Figure 1) but differed in that they were intended for a single user, and would be meaningful to an experimental participant within the context of a 1-hour duration interaction (for long-term studies a duration of 1 hour maximum for each session was considered appropriate in order to avoid fatiguing the participants). The scenarios were grounded in an imagined daily life, with the robot adopting an assistive role: allowing the participants to inform the robot about their preferences in terms of drinks, snacks, leisure activities, and TV programs that they preferred. These elements were used in individual episodes whereby each scenario was performed twice during the long-term studies, according to the schedule shown in Table 1. In these episodes, the participants were asked to engage in a structured role-play-like scenario (Seland 2009) in order to investigate the role of the robot in a manner that could be directly related to the participants’ everyday experience.

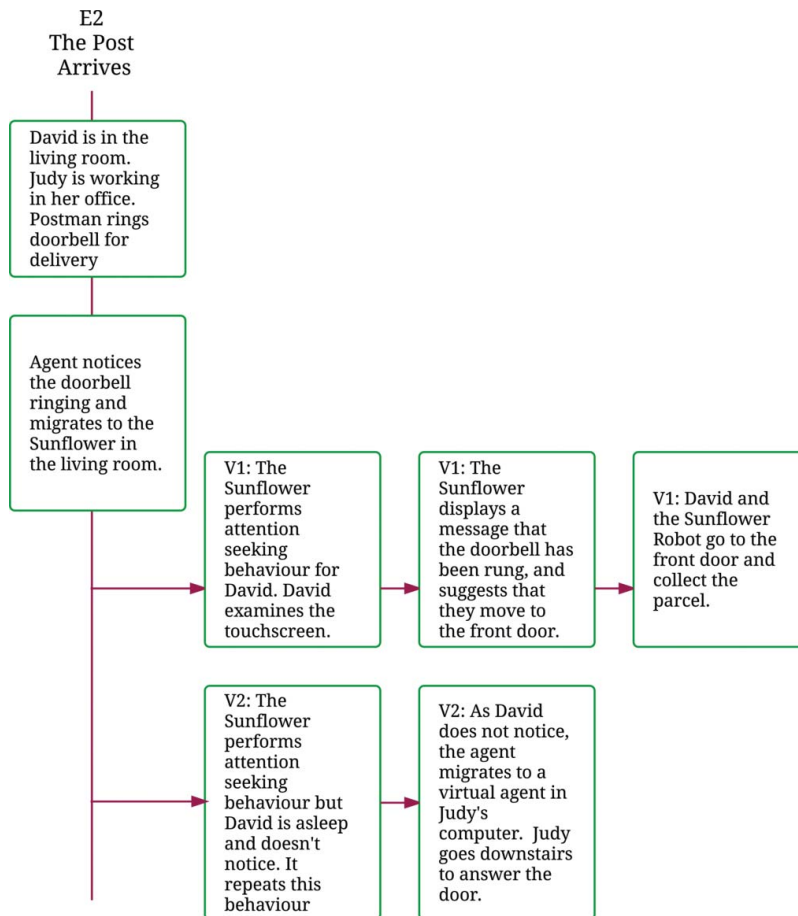


FIG. 2. Episode from a “Normal Day” for the user personas (2).

TABLE 1
Overview of sessions

Week	Session content
Week 1	Introduction to the Robot House, familiarization with the robots and their interface. Baseline experiment.
Week 2	Review of Robot House, robots, and interface. Repeat of experiment.
Week 3	Open-ended scenario A
Week 4	Open-ended scenario B
Week 5	Repeat of experiment
Week 6	Open-ended scenario A
Week 7	Open-ended scenario B
Week 8	Repeat of experiment
Week 10	Debriefing

Note. The constrained experiment was run in Week 1 (2 tasks, Human-only condition) and Weeks 2, 5, and 8 (two tasks, Human-Only and Robot and Human conditions).

Therefore, they could directly experience the impact of the robot. These scenarios also investigated particular issues that were of interest to our research, such as human and robot communication and “agent migration” (see explanation in the following).

These scenarios were based around episodes in two “imaginary” days and were intended to investigate interactions with and responses to the robot in an everyday setting. The first episode took place in the “morning” and focused on the expressive capabilities of the Sunflower robot. The second episode was set during the “afternoon” and was focused on the participants’ impression of agent migration—the ability of an agent’s “mind” to move between different robot and virtual embodiments (Syrdal et al. 2009; Duffy et al. 2003). Here, the agent’s “mind” comprises its memory, its interaction history, and a sense of context; for example, it can remember the user’s preferences while moving between different embodiments, and can continue tasks begun in one embodiment within another. This allows the agent to take advantage of features and functionalities of more than one embodiment while maintaining the persistent features that make it unique and recognizable from a user’s perspective. These attributes include awareness of interaction history and context, as well as persistent customizable features. In the scenario, the migration took place between a Sunflower and a SONY Aibo robot. For both of these scenarios, participants were briefed as to the time of day and the particulars of the situation they were going to take part in (Koay et al. 2011).

CONSTRAINED EXPERIMENTS

Cognitive Prosthetic

The scenarios identified several instances in which the robot companion would be able to assist the user by providing information. This information could be provided in the form of reminders of appointments, mealtimes, and medicines. In the chosen scenario the robot’s task was to remind “David” to take his heart medication.

Adherence to a prescribed regimen of medication can be difficult for many patients. Early approaches (as exemplified by Schwartz et al. 1962) presented this as being caused by a shortfall in the ability of the patient, who was seen as making mistakes. More recent approaches consider a wider range of reasons for nonadherence to prescribed medicine regimens. In addition to the cognitive abilities of the patient, the new approaches also take into account other factors such as the complexity of the medication schedule, perceived efficacy of the treatment, and perceived risk of side effects (Horne et al. 2005).

While this particular scenario used the robot purely to remind the user of his schedule in a manner similar to that of cognitive prosthetics on hand-held platforms (Modayil et al. 2008), this functionality can also be combined with more persuasive technologies that use relational and other strategies in order to encourage habits conducive to the health of the user (Bickmore et al. 2005). However, this was not the focus of the current study, which focused purely on the cognitive prosthetic aspect of such technologies and its impact within the performance of a task.

The experimental instantiation of the Cognitive Prosthetic task involved participants putting Scrabble tiles into the correct spaces of a medicine dispenser on the living room table (see Figure 5, shown later), relying on a master list that had to remain on the kitchen bench. There were 28 spaces for the tiles, and both the position of the tiles in the dispenser and their position on the list in the kitchen were randomized.

Fetch and Carry

The Fetch and Carry task involved the carrying of objects between different rooms. This task was performed during episodes such as mealtimes, where the robot could assist with the movement of prepared food from the kitchen to the dining area and returning of dishes to the kitchen. It was also considered to be of utility in the episodes where “David” could use it while engaging in his hobby, for example, to move models and tools from storage to a work surface in a different room.

The term Fetch and Carry comes from Hüttenrauch and Eklundh (2002), who in their case study describe how a user with partial mobility impairment uses a

mobile robot as a platform for transporting objects that this person would otherwise be unable to move without assistance from another person. This particular task is interesting due to both the utility of the task and the human-robot interaction issues that it highlights.

The Fetch and Carry capability of robots can be of use to a wide variety of users because there are many reasons why they may need assistance for transporting objects, ranging from fall injuries to neurodegenerative conditions like Parkinson's (Kamsma et al. 1995; Walker and Howland 1991). It is also an interesting task from a human-robot interaction perspective, as it is unique to the physical nature of robots and involves both human and robot interactants negotiating and moving in a shared physical space. As long as the robot is capable of moving between two or more points and is fitted with a suitable container for the transport of objects, a robust and stable realization of this task is well within the current state of the art. For a product prototype implementation for this task, see the Danish Technological Institute (DTI) robot-butler "James" (Danish Technological Institute 2012).

The experimental instantiation of the Fetch and Carry task involved the participants moving 100 plastic balls from a net on the kitchen bench to the living room table using only one hand. This was a constraint that was easily implemented while being challenging to the participants. While the balls were very light, requiring little physical strength, they were quite unwieldy in numbers larger than four or five, so required several trips back and forth to transport them all.

Assistance as envisaged with the Cognitive Prosthetic and Fetch and Carry tasks can be used in response to changed circumstances, such as recovery from illness and accidents, as well as rehabilitation after strokes, where the prospective user will have to learn new skills to aid in daily living, or gradually recover mastery of old skills.

For the experimental instantiation of both these tasks, we decided to choose tasks that, while not strenuous, would present a challenge to the participants, and in which the use of a robot would have a clear impact on the task. In addition, it was hoped that the experimental constraints would add novelty to the task, allowing us to see the impact of changes in participant task mastery.

RESEARCH QUESTIONS

Differentiation of Tasks on the NASA TLX

The first research question was whether or not we could differentiate between the tasks using their NASA Task Load Index (TLX), a measure for different types of workload that is described in more detail in the methodology section. It was expected that the two tasks would load more strongly on their "primary" dimensions (namely,

Fetch and Carry along the Physical Dimension, and Cognitive Prosthetic along the Mental Dimension). It was also of interest to see whether these tasks changed over time (i.e., whether practice changed the nature of the tasks in terms of experienced workload).

Research Question 1:

- (a) How did the two tasks differ from each other in terms of experienced workload at the initial presentation?
- (b) How did the experience of the tasks change over time?

Impact of the Robot

We were also interested in how the use of the robot would alter the perceived workload of the two tasks, and how this impact changed over time. While we expected the use of the robot to impact the different tasks along their primary dimensions by reducing participants' workload in the initial interactions with the robot as an aid, we were also interested in how the robot impacted the workload on these tasks along the other dimensions.

Research Question 2:

- (a) How did the robot impact the experienced workload on the tasks along the different dimensions of the NASA TLX?
- (b) How did the impact of the robot change over time?

The Experience of the Task and the Robot

Our final interest was in how participants reasoned about the tasks, and how they described the tasks in terms of what contributed to their workload and their experience of the robot's assistance.

Research Question 3:

- (a) How did the participants reason about the tasks? Did they see them as "natural" and relevant to their own everyday experience?
- (b) How did participants describe the role of the robot in the task? What were the benefits of its use, and what were the drawbacks?

METHODOLOGY

Apparatus

Two robots were used in this study. The first was the UH Sunflower robot, which uses a Pioneer base (commercially available from MobileRobots) but with significant modifications (See Figure 3). The main mode of direct interaction with this robot is its touch-screen (Figure 4), which can be used to both display information to the user and issue commands to the robot. Sunflower also has an extending tray that can be used to carry light objects.

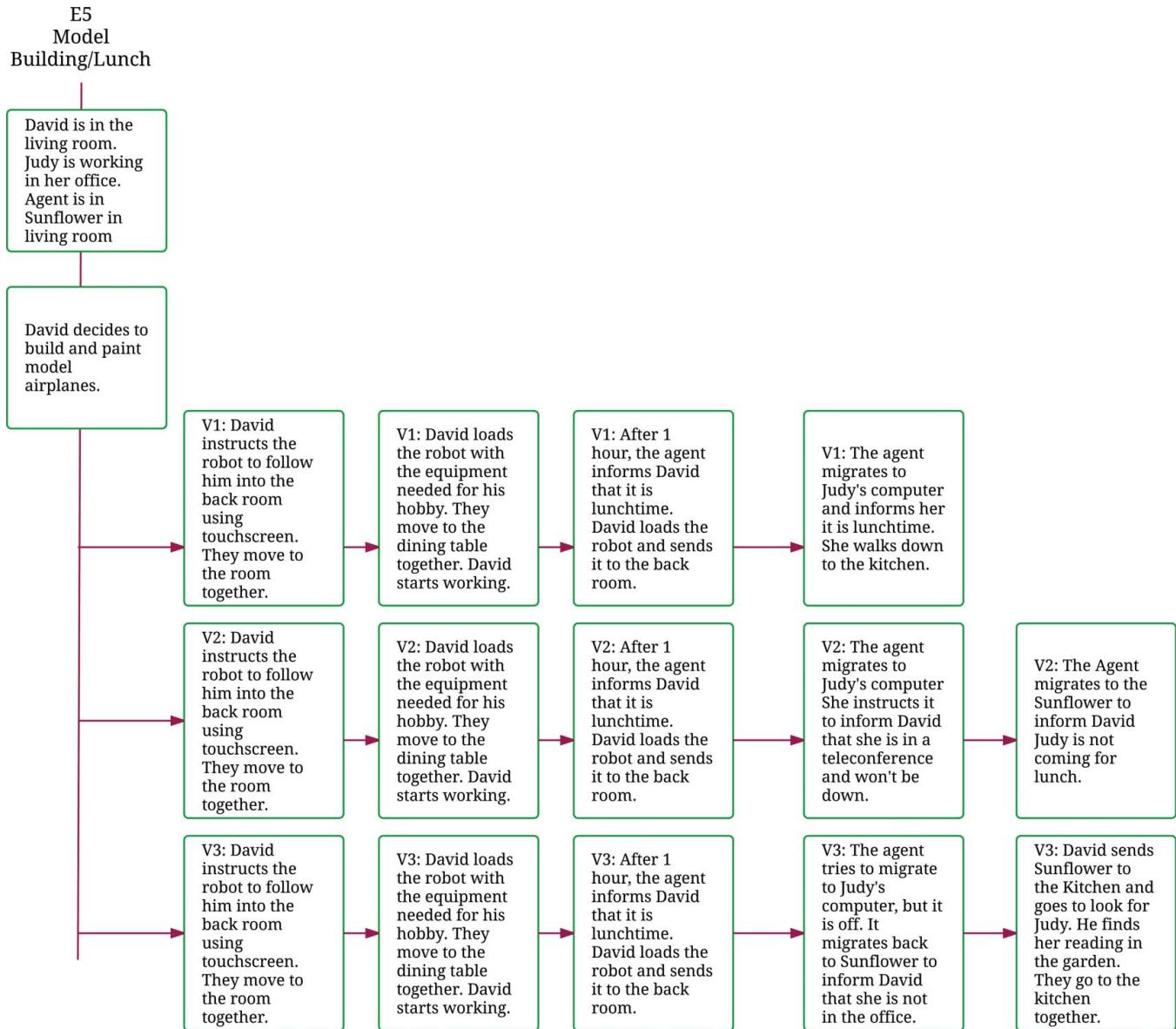


FIG. 3. The Sunflower robot used in this study. The robot was built at the University of Hertfordshire, significantly extending a basic Pioneer Platform.

The Sunflower robot is similar in shape and interaction capabilities to other robots intended for domestic use (e.g., Coradeschi et al. 2013; Lammer et al. 2014; Koay et al. 2014). The second robot used in the study was a SONY AIBO.¹

In addition, laptop PCs were set up for Skype calls.

The apparatus for the Fetch and Carry task consisted of the previously mentioned 100 play balls. The apparatus for the Cognitive Prosthetic task was comprised of the generic medicine tray and scrabble tiles as shown in Figure 5. Both of these are widely available commercially.

Experimental Setup

Participants were asked to visit the robot house once a week for a period of 10 weeks, in order to study how participants' views of, and interactions with, the robots changed over time. See Table 1 for an overview of the sessions that the participants took part in. References made in this article to a specific week are based on Table 1. While the participants would only do the controlled, task-based prototyping experiment in weeks 1, 2, 5, and 8, it is important to note that in these other sessions they interacted with the robot, using its touch-screen



FIG. 4. Interacting with the touch-screen interface on the Sunflower robot.

interface and moving in the same space as the robot, thus familiarizing themselves with the robot and its use between the constrained task-based experiments. Each session took about 1 hour, including debriefing.

Procedure

Introduction. The introduction session introduced the UH Robot House and the robots to the participants. The participants were instructed in the use of the Sunflower robot and touch-screen, as well as how this robot responded to scheduled and sensor events. The participants were given a tour of the living areas where they would interact with the robot, and were shown the kitchen cupboards and fridge shelves that would be “theirs.” In addition, they were introduced to the AIBO robot and its use in remote human-human interaction scenarios. Throughout this tour, participants were encouraged to think of these areas as their home and to put themselves in the mind set of someone living in the house. This was intended to begin the process of framing the narrative (Dindler and Iversen 2007) of the open-ended scenarios. It was also intended as a session in which the participants could make themselves as comfortable in the house as possible. The session ended with the baseline experiment.

Open-ended scenarios. As mentioned earlier, there were two open-ended scenarios that were presented twice to the participants. At the beginning of each open-ended scenario session, the participants were given a narrative framing of the context of the scenario that they were taking part in. They were told the time of day, and also what had transpired immediately before the beginning of the scenario. Scenario A began in the morning and the participants were told the following:

Imagine that you have now woken up. In the introductory session you gave us some preferences for what you would like to do in the early morning. The robot has stored these preferences and will try to help you do them. When you are ready, you will come out of the bedroom and sit down on the sofa. The robot will then approach you.



FIG. 5. Medicine dispenser and Scrabble tiles used in the Cognitive Prosthetic task as part of the controlled experiments.

Scenario B began in the afternoon:

Imagine that it is afternoon and you have just returned home and have just sat down on the sofa. You have planned to watch some TV. In the introductory session, you gave us some preferences as to what TV programs you like to watch and also what sorts of snacks and drinks that you prefer to eat. The robot has stored these preferences. It will also respond to events such as phone calls and doorbells. When you are ready to begin, sit down on the sofa and the robot will approach you.

After this briefing, the scenarios ran as outlined previously. Participants were asked to fill in questionnaires after the scenario was completed.

Constrained experiments. There were two sets of conditions for the experiment:

1. Task:
 - a. Fetch and Carry.
 - b. Cognitive Prosthetic.
2. Robot:
 - a. Human-Only.
 - b. Robot and Human.

In the baseline experiment in Week 1, participants undertook both task conditions in the human-only condition. The presentation order of the two tasks was counter-balanced in order to account for a presentation effect. In weeks 2, 5 and 8, participants did both task conditions for both of the robot conditions for a total of 4 trials in these weeks. The presentation order was within each week, for both Task and Robot conditions. Participants were given a questionnaire to respond to after each run of a task.

Robot Use

The use of the robot was adapted to each task: For the Fetch and Carry task, participants were allowed to use the extendible tray of the robot as an additional platform to transport the plastic balls to the living room table. The participants could instruct the robot to move between the locations using the touch-screen interface. For the Cognitive Prosthetic task, the participants could access the list through the touch-screen interface. The participants could only access one quarter of the list at any given time, and could only choose which portion of the list to access while in the kitchen. This meant that in order to access the whole list, they would have to make several journeys between the living room and the kitchen over the course of the trial.

Instructions. Before each task, participants were shown the apparatus involved in each task, and had the task explained to them. For the robot condition, participants were shown how to use the robot, and how to operate the touch-screen interface relevant for that particular task. Participants were asked to try to complete the task as quickly as possible. They were told that their performance was not being assessed, and that if the task took longer than 10 minutes to complete, the experimenters would stop the experiment.

Measures: NASA Task Load Index

We used the NASA Task Load Index (TLX) as the primary measure for the evaluation of the constrained tasks. The NASA TLX is a questionnaire-based means of measuring workload for specific tasks along several different dimensions. It is particularly intended for examining human-machine interactions (Hart and Staveland 1988). As it is a posttask measure, administering it to a participant would not affect task performance in the manner that a concurrent measure such as a think-aloud protocol might (Russo et al. 1989). Despite it being a subjective, posttask measure, studies have shown it to be a reliable and valid tool for examining task difficulty and performance (Rubio et al. 2004). Since its conception, it has

TABLE 2
Dimensions of the NASA Task Load Index

Dimension	Workload in terms of . . .
Mental	. . . reasoning remembering, planning, thinking
Physical	. . . strength and endurance, dexterity
Temporal	. . . pace, time pressure, speed
Performance	. . . success and satisfaction
Effort	. . . effort needed to accomplish performance
Frustration	. . . annoyance, frustration, stress

TABLE 3
Open-ended questions

Q1. What was the most difficult part of doing the task?
Q2. What would have made the task easier?
Q3. What were the benefits of doing the task with the robot?
Q4. What were the drawbacks of doing the task with the robot?

been used across a wide variety of domains and tasks (Hart 2006). It was chosen over the more focused Human-Robot Interaction Workload Measurement (HRI-WM) (Yagoda 2010) because the main focus of our study was on the participants' experience of the tasks themselves, rather than an assessment of how they interacted with the robot. The NASA TLX measures workload along six dimensions, shown in Table 2.

Ad Hoc Questions

In addition to the NASA TLX, participants were asked open-ended questions, inviting them to describe their experiences of the tasks themselves, as well as the role of the robot within them. These questions are shown in Table 3.

Participants

Twelve participants took part in the study, recruited through advertisements on the University of Hertfordshire Intranet, mailing lists, and social networks. There were eight males and four females in the sample. The mean age was 32 years and the median age was 26 years, and the age range was 18–64 years. The use of human participants had been approved by the University of Hertfordshire Ethics Committee under protocol number 1112/39.

RESULTS

The results for the constrained tasks with respect to the original research questions were as follows.

Research Question 1: Characteristics of the Task

Baseline values. The differences between the two tasks were examined using a series of t-tests (Table 4 and Figure 6). As could be expected, the TLX significantly differentiates between the two tasks in terms of the physical and mental dimensions. The most salient differences between the two can be seen along the Physical Dimension (which contributes significantly more to the workload of the Fetch and Carry task) and the Mental

TABLE 4
TLX baseline scores for tasks

Dimension	Fetch and Carry mean (SE)	Cognitive Prosthetic mean (SE)	Mean difference	95% CI	<i>t</i> (df)	<i>p</i>
Mental	0.50 (.14)	2.75 (.58)	-2.25	-3.44 to 1.07	4.20 (11)	** .01
Physical	2.46 (.62)	0.70 (.25)	1.75	0.32 to 3.19	2.69 (11)	** .02
Temporal	1.89 (.51)	1.95 (.46)	-0.07	-0.99 to 0.86	-0.16 (11)	.88
Performance	0.83 (.33)	0.63 (.26)	0.19	-0.72 to 0.86	0.46 (11)	.65
Effort	1.15 (.30)	1.40 (.30)	-0.25	-0.72 to 1.10	-0.62 (11)	.55
Frustration	1.51 (.47)	0.58 (.18)	0.94	-0.20 to 2.07	1.82 (11)	.10

Dimension (which contributes significantly more to the workload of the Cognitive Prosthetic task). There is a trend approaching significance for the Frustration Dimension, which suggests that it contributes more to the workload of the Fetch and Carry task.

Long-term change. Change across the 8 weeks for the Fetch and Carry task is described in Table 5 and Figure 7. They suggest that the only significant change for this task was along the Effort Dimension, which contributed more to the workload in this task in later weeks than the first week. Change across the 8 weeks for the Cognitive Prosthetic task is described in Table 6 and Figure 8, suggesting that overall there were no significant changes for this task in terms of what dimensions contributed to the workload on this task. However, a trend approaching significance indicates that the Temporal Dimension contributed less to the workload of this task in later weeks. Also, while the descriptive statistics of Table 6 suggest that there was an equally substantial mean change in the Mental Dimension, the variance between participants' individual scores stopped this change from being significant for this sample.

Research Question 2: Robot Impact

Fetch and Carry. The overall impact of the robot can be found in Table 7 and Figure 9. There were significant main effects for the role of the Robot along the Physical, Temporal, Performance, and Effort dimensions. However, all of these main effects, with the exception of Performance, were mediated by interaction effects between the role of the robot and the long-term effects, so we consider these interaction effects in the text as well. For Performance, there was a main effect for robot assistance. This effect suggests that performance was experienced as worse with the robot than if the participant acted on his or her own. This effect was very pronounced in week 2 but decreased with time. For the Physical dimension, there was a significant interaction effect between time and assistance. The relationship suggested by the descriptive statistics in Table 7 and Figure 9b is that the participants found that the robot reduced the workload overall but this effect decreased after week 2. For the Temporal Dimension, there was a significant main effect described in Figure 9c, where participants found that the robot overall increased the Temporal aspects of workload. The interaction effect approaching significance,

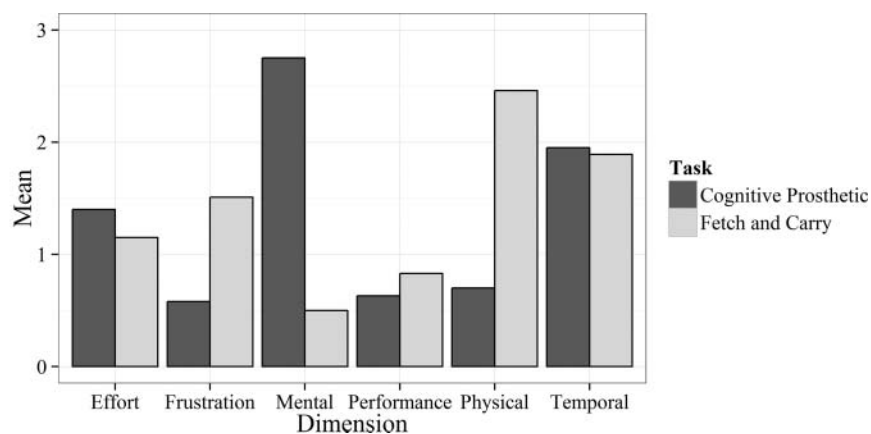


FIG. 6. TLX baseline scores for tasks.

TABLE 5
Long-term change for Fetch and Carry task

Dimension	Week 1	Week 2	Week 5	Week 8	$F(3, 8)$	p	η^2
Mental	0.50 (.14)	0.19 (.06)	0.20 (.06)	0.17 (.14)	0.99	.45	.27
Physical	2.46 (.62)	3.01 (.46)	3.09 (.58)	2.70 (.56)	2.44	.14	.48
Temporal	1.89 (.51)	1.30 (.29)	2.25 (.44)	1.95 (.50)	1.20	.37	.31
Performance	0.83 (.33)	0.58 (.19)	0.83 (.34)	0.34 (.05)	2.15	.17	.45
Effort	1.15 (.30)	2.50 (.37)	1.84 (.42)	2.45 (.40)	5.48	*.02	.67
Frustration	1.51 (.47)	0.95 (.25)	0.88 (.20)	1.02 (.34)	1.52	.28	.36

however, suggests that this effect decreased over time. The robot's impact on the Effort Dimension was quite small in weeks 2 and 5. However, by week 8, the assistance of the robot reduced the workload along this dimension (see Figure 9e).

Cognitive Prosthetic. The overall impact for the robot on the Cognitive Prosthetic Task is shown in Table 8 and Figure 10. The impact of robot assistance was primarily along the Mental, Performance, and Effort dimensions. There were no interaction effects.

Participants viewed the robot as reducing workload along the Mental Dimension. This was consistent across the 3 weeks. On the other hand, the descriptive statistics in Table 8 suggest that participants saw the robot as adding significantly to the workload along the Performance Dimension (i.e., making it harder to succeed on the task). This effect is less pronounced in the last week. The other significant impact was along the Effort Dimension. The descriptive statistics in Table 8 suggest that participants found they needed to exert less effort when aided by the robot. There were also two nonsignificant trends for the Temporal and Frustration dimensions. These trends suggested that the participants saw the use of the robot as contributing to more workload in these two dimensions,

thus making the task both more frustrating and time-critical.

Research Question 3: The Experience of the Task and the Robot's Role

The analysis of participant responses to qualitative questions (see Table 3) was conducted in two main stages. In the first stage, one of the researchers examined the open-ended qualitative responses from the questionnaires and categorized them into primary themes and subthemes for each task and each week. These themes were then examined across weeks for each of the tasks. This led to the collection of themes identified in the first two columns in Tables 9 and 10. A unified category scheme for both tasks could not be developed, largely due to the large qualitative differences between the tasks. After this categorization, two of the researchers went through the responses and categorized them as major (+), minor (−), and nonexistent (0).

The themes that were the most prevalent in the responses were categorized as major. Minor themes were those less prevalent but still reported by a small group of participants. Themes that did not appear in the responses for a particular week were categorized as nonexistent.

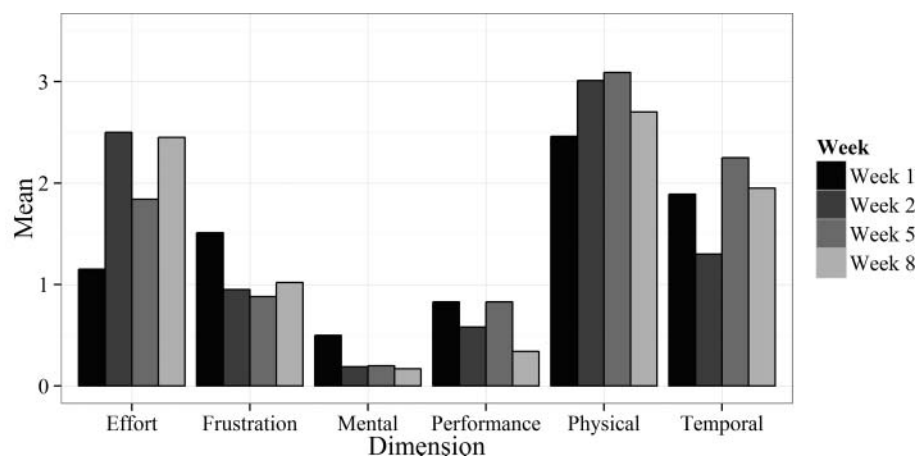


FIG. 7. Long-term change for Fetch and Carry task.

TABLE 6
Long-term change for Cognitive Prosthetic task

Dimension	Week 1	Week 2	Week 5	Week 8	$F(3, 8)$	p	η^2
Mental	2.75 (.58)	1.97 (.48)	2.21 (.51)	1.81 (.38)	1.29	.34	.33
Physical	0.70 (.25)	0.72 (.30)	0.63 (.26)	0.60 (.18)	0.23	.87	.08
Temporal	1.95 (.46)	1.78 (.44)	1.04 (.30)	1.22 (.43)	3.87	*.06	.59
Performance	0.63 (.26)	0.93 (.28)	0.76 (.22)	0.68 (.32)	1.63	.26	.38
Effort	1.40 (.30)	1.32 (.31)	1.25 (.27)	1.22 (.28)	1.23	.36	.32
Frustration	0.58 (.18)	1.08 (.42)	0.60 (.21)	.32 (.10)	2.31	.15	.46

The final categorization and assignment of the themes was done by the researchers, after having compared their coding of responses, discussed discrepancies, and reached a consensus.

Fetch and Carry. The themes emerging from the participants' responses are described in detail next and summarized in Table 9:

Week 1. For the Fetch and Carry task, the two primary themes emerging from Q1 (What made the task difficult?) were the physical difficulty of handling the balls and the constraint of using only one hand when performing the task. They were also evident in the responses to Q2 (What would have made the task easier?) where the possibility of release from this constraint was the predominant theme.

Week 2. Week 2 saw the introduction of the robot, and Q1 and Q2 were asked for both the human-only and the robot-human condition. For the human-only condition, the theme of the constraint continued among some of the participants. Participants would also contrast the human-only condition with the use of the robot when answering questions related to both conditions. When contrasting the conditions, participants highlighted the practical benefit of being able to perform the tasks in fewer trips. However, the second most prevalent theme in the

participants' statements was the slow speed of the robot. The sample as a whole agreed that the speed of the robot was problematic from a task perspective, with participants having to change their speed of performing the task to accommodate the robot. This was achieved either by walking (more slowly) with the robot to the living room and back, or by waiting at the appropriate place to load or unload to the robot. In response to Q2 for the robot condition, the participants overwhelmingly suggested increasing the speed of the robot and/or the size of the tray. They also suggested that an ability of the robot to manipulate objects by loading itself would be helpful. In addition to purely task-related comparisons, a small group of participants highlighted interactional aspects of doing the task with the robot: that the robot provided company or that the task was more enjoyable when using the robot.

Week 5. Week 5 saw a continuation of the same themes as in week 2. New themes also emerged related to how participants rated their own performance. Some participants identified changes in their own behavior between conditions. They referred to a type of *social loafing* (Latane et al. 1979) that occurred when they did the task with the robot, and that they put more effort in when they were doing the task by themselves. Other

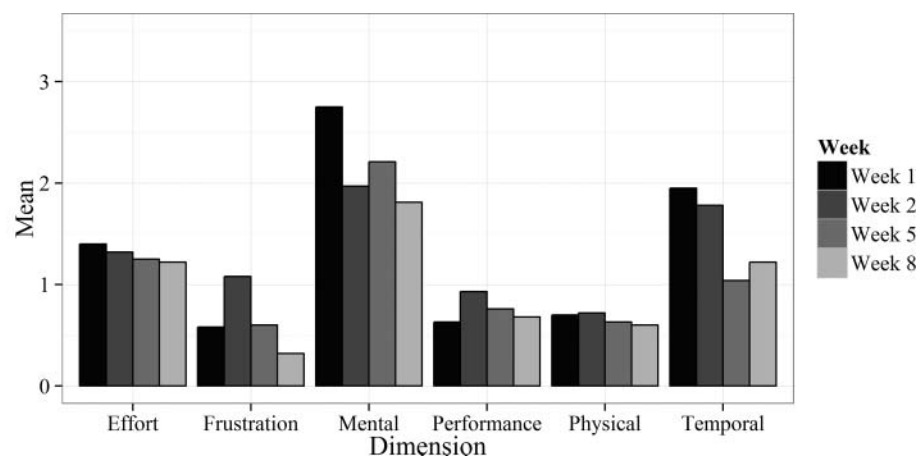


FIG. 8. Long-term change for Fetch and Carry task.

TABLE 7
Robot impact on Fetch and Carry

Dimension		Week 2	Week 5	Week 8	ME	ME	ME	In F	In	In
					F					
Mental	Human	0.20 (0.20)	0.20 (0.20)	0.17 (0.25)	3.83	.08	.29	1.24	.33	.22
	Robot	0.95 (1.54)	0.27 (0.20)	0.20 (0.20)						
Physical	Human	3.26 (1.41)	3.09 (1.91)	2.70 (1.86)	16.16	.01*	.62	13.17	.01*	.75
	Robot	0.26 (0.54)	1.83 (1.29)	1.70 (1.07)						
Temporal	Human	1.39 (1.01)	2.25 (1.47)	1.95 (1.67)	7.65	.04*	.35	4.00	.05*	.47
	Robot	6.10 (5.84)	2.25 (1.68)	1.67 (1.47)						
Performance	Human	0.61 (0.67)	0.83 (1.13)	0.33 (0.16)	7.65	.02	.43	1.41	.29	.24
	Robot	2.01 (3.11)	1.46 (1.78)	0.56 (0.89)						
Effort	Human	2.72 (1.03)	1.84 (1.40)	2.46 (1.33)	4.61	.05*	.32	5.64	.03*	.56
	Robot	2.28 (2.50)	1.61 (1.00)	1.22 (0.96)						
Frustration	Human	1.03 (0.85)	0.88 (0.66)	1.02 (1.14)	1.83	.21	.16	2.28	.16	.34
	Robot	0.21 (0.19)	0.65 (0.96)	1.05 (0.85)						

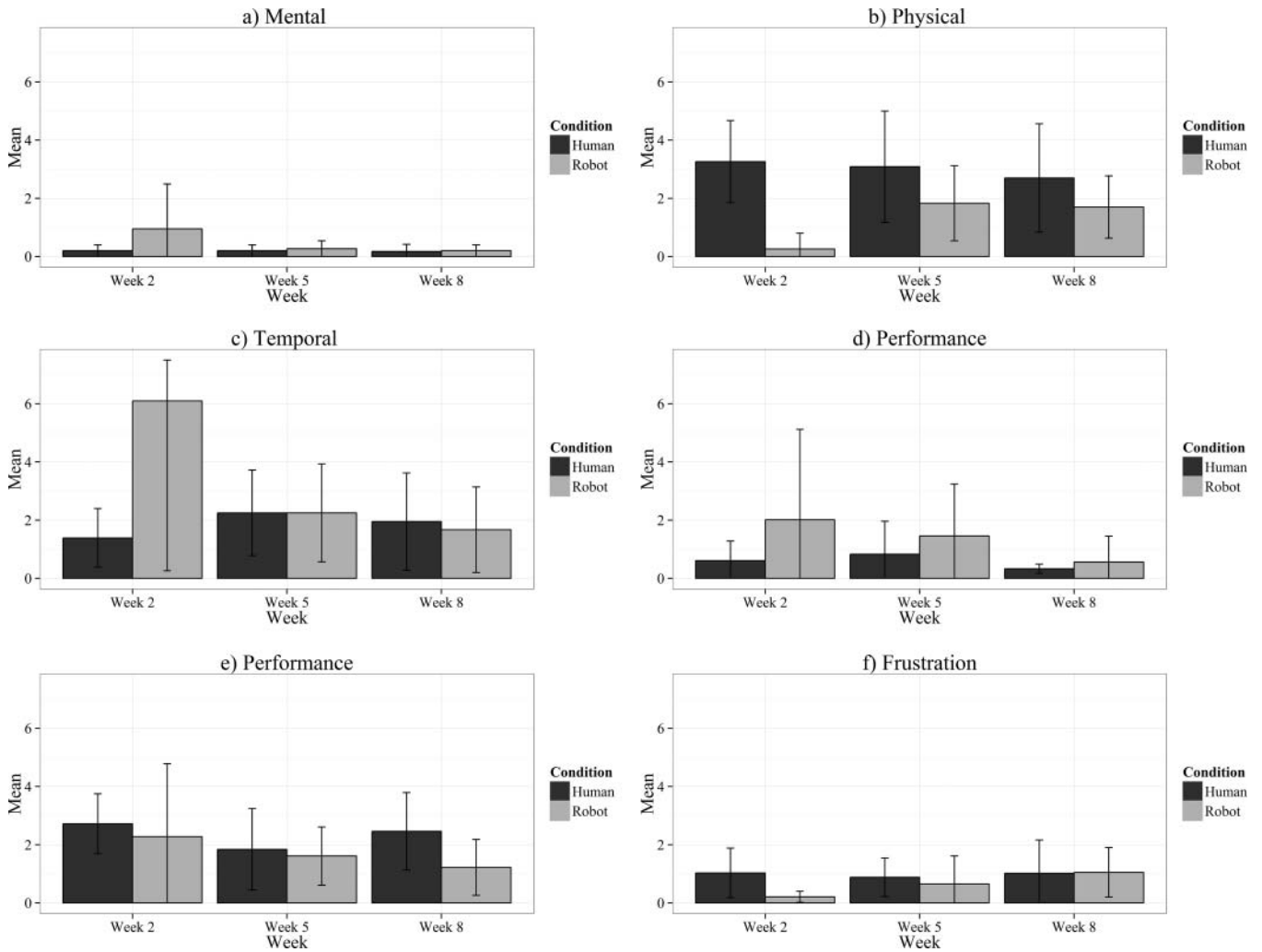


FIG. 9. Robot impact on Fetch and Carry in terms of experienced workload.

TABLE 8
Robot impact on Cognitive Prosthetic

Dimension		Week 2	Week 5	Week 8	ME <i>F</i> (3, 8)	ME <i>p</i>	ME η^2	In <i>F</i> (3, 8)	In <i>p</i>	In η^2
Mental	Human	2.13 (1.63)	2.21 (1.71)	1.81 (1.26)	16.24	.01*	.62	0.42	.67	.08
	Robot	0.76 (0.90)	0.73 (0.98)	0.88 (1.72)						
Physical	Human	0.77 (1.08)	0.63 (0.85)	0.60 (0.61)	0.34	.58	.01	2.80	.11	.38
	Robot	0.36 (0.38)	0.49 (0.65)	0.83 (0.75)						
Temporal	Human	1.93 (1.53)	1.04 (0.99)	1.22 (1.41)	3.63	.09	.27	0.48	.64	.10
	Robot	2.38 (1.64)	1.92 (1.63)	1.04 (0.99)						
Performance	Human	0.99 (0.99)	0.76 (0.72)	0.68 (1.05)	5.90	.04*	.37	1.23	.34	.21
	Robot	2.19 (1.14)	1.14 (1.51)	0.81 (1.06)						
Effort	Human	1.41 (1.05)	1.25 (0.90)	1.22 (0.94)	4.79	.05*	.32	0.23	.8	.05
	Robot	0.94 (0.77)	0.72 (0.87)	0.95 (0.70)						
Frustration	Human	1.18 (1.47)	0.60 (0.69)	0.32 (0.32)	3.21	.10	.23	0.56	.59	.11
	Robot	1.90 (2.16)	1.10 (1.47)	1.34 (1.33)						

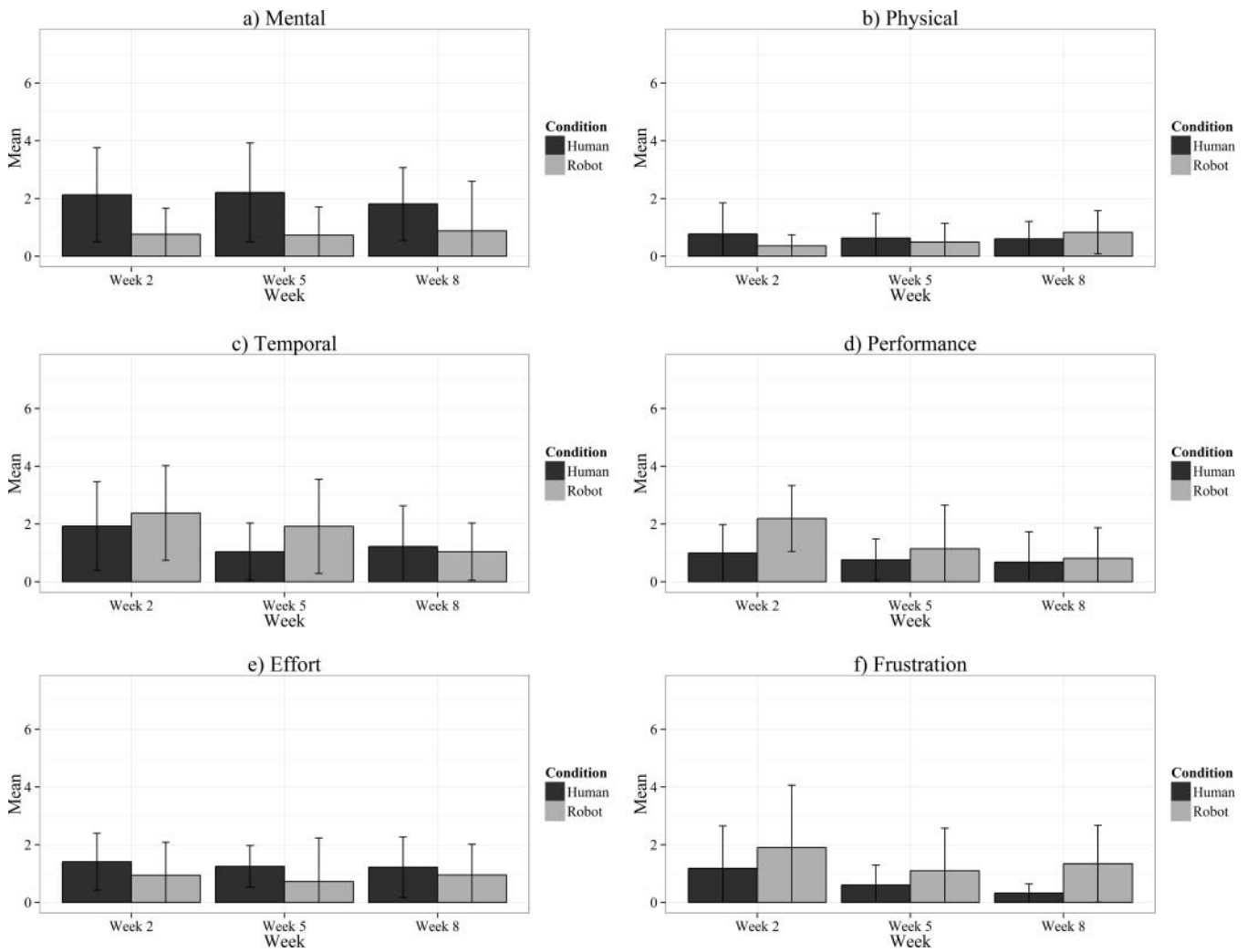


FIG. 10. Robot impact on Cognitive Prosthetic in terms of experienced workload.

TABLE 9
Themes for the Fetch and Carry task

Primary theme	Subtheme	Week 1	Week 2	Week 5	Week 8
Imposed constraint	Using one hand	+	+	+	-
	Benefit from the tray	0	+	+	+
Use of the robot	Mutual adaptation	0	0	-	+
	Interface	0	0	+	+
	Having to wait	0	+	+	-
Speed of the robot	Walking with the robot	0	+	0	0
	Changing speed	0	+	+	+
Changing capabilities	Changing tray	0	+	+	+
	Object manipulation	0	+	0	-
	Robot as partner	0	+	+	+
Interactional aspects	Enjoyment	0	+	+	+
	Social loafing	0	0	+	+

Note. + Theme present; - theme present to a lesser degree than in the other weeks; 0 theme not present.

participants highlighted mutual adaptation. They reported that they were getting better at coordinating their own and the robot's roles in the task, reducing waiting, and making the use of the robot more efficient. The most common strategy was to perform the task in an asynchronous manner, only loading and unloading the robot at convenient times instead of synchronizing each trip. However, for the sample as a whole, the theme of having to wait for the robot was still prevalent. In addition, this week saw statements regarding the touch-screen interface for this task. There were no statements regarding object manipulation capabilities in this week.

In addition, participants continued to reference the social aspects of doing the task with the robot.

Week 8. Week 8 was very similar in terms of themes to Week 5. The main difference was one of prevalence. The theme of mutual adaptation continued and was more widespread, while the theme of having to wait for the robot was much less prevalent this week.

Cognitive Prosthetic. The themes arising from the participants' responses for this dimension are described below and summarized in Table 10.

Week 1. In week 1, two main themes arose in participant responses to Q1. The first was the difficulty of

TABLE 10
Themes for the Cognitive Prosthetic task

Primary theme	Subtheme	Week 1	Week 2	Week 5	Week 8
Imposed constraints	Separation of list and dispenser	+	-	-	-
	Robot positioning	0	-	+	+
	Random order of tiles and position in list	-	+	-	-
Performing the task	Difficulty in trying to remember	+	-	-	-
	Physically manipulating the tiles	0	0	-	+
	Use of strategy	-	-	+	+
Nonrobotic tool	Pen and paper	+	-	-	-
	Tray	+	+	-	-
Robot benefits	Easy	0	+	+	-
	Infallible/no pressure	0	+	+	+
	Subversion	0	0	+	+
	Slow	0	+	+	+
Robot drawbacks	Flexibility	0	+	+	+
	Interface issues	0	-	+	+
	Control	0	+	+	+

Note. + Theme present; - theme present to a lesser degree than in the other weeks; 0 theme not present.

having to remember the position of the tiles while walking from the kitchen to the living room. The second was the attempt at developing a strategy for solving the task without having to rely on memory alone. Responses to Q2 did, as for the Fetch and Carry task, focus on the constraints of the task—in particular, the placement of the list of tile positions in a separate location from the medicine dispenser, and the list of tiles not being in any discernible order. A small group of participants managed to develop a strategy for doing this task more efficiently, which consisted of arranging tiles spatially in one's palm in the same manner that they were to be arranged in the medicine dispenser and then transporting them over and inserting them into the dispenser in the same order. The final theme was an expressed desire for tools to aid in the task. There were two categories of tools: reminder tools, such as a pencil and paper to jot down the appropriate tiles and their positions, and tools to make the strategy described earlier more efficient. An example of the latter would be a large tray to arrange and carry all the tiles on at once.

Week 2. In the human-only condition, the adoption of the strategy just described became more prevalent as fewer participants relied on memory alone to perform the task. This change was also reflected in the suggestions for tools to be used, where items that would aid in the use of this strategy were suggested to a larger extent than in the previous week. When discussing the role of the robot, participants raised several issues. They considered the robot-assisted solution of the task to be easier, as there was no need to either remember anything or adopt a strategy. Participants in particular referred to the infallibility of the robot's memory and how this made them feel less under pressure to perform the task correctly.

However, participants referenced the interaction with the robot in itself as a source of difficulty for the task as well. The robot was also described as slow and lacking in flexibility. The relinquishing of control to the robot was also referenced when discussing the procedure used to access the information on the robot.

Week 5. The results in week 5 followed many of the same themes as week 2. There was a continued increase in the use of the strategy outlined in week 1. By this week the majority of participants used this strategy for the human-only condition. References to interface issues were more prevalent in this week's responses, as were references to the physical aspect of the task, such as manipulating and putting the tiles in the dispenser. This week also saw a new theme of subversion emerging. Two of the participants described how they used the robot the way they wanted to, instead of how they felt they were being expected to. They arranged the tiles spatially on the tray of the robot in the kitchen and then used it to transport them in the correct arrangement to the

dispenser in the living room, thus sidestepping the use of the robot as a Cognitive Prosthetic.

Week 8. Week 8 results were similar to those in week 5. Statements related to the physical carrying out of the task were more prevalent this week than on any other week. The majority of participants stated that the task had become easier for them to do. However, many still referenced the benefits of the robot, in particular its infallibility.

DISCUSSION

Research Question 1—Differences Between the Tasks

We were able to differentiate between the tasks in terms of their NASA TLX profile. Initially, the two tasks were significantly different from each other only along their primary dimension, with a trend for the Fetch and Carry task loading more on the Frustration Dimension. In terms of long-term change, however, the picture was slightly different. While neither of the two tasks changed on the Frustration Dimension, they did change along other dimensions. The Fetch and Carry task changed in terms of Effort, and loaded higher on this dimension in the later weeks. The Cognitive Prosthetic task changed along the Temporal Dimension, and time pressure was considered less important in weeks 5 and 8. This suggests that the use of the NASA TLX for HRI tasks in domestic environments was a valid and meaningful approach.

Research Question 2—Impact of the Robot

The robot changed the participants' experience of the two tasks differently, both in its initial use as well as over time. For the Fetch and Carry task, the robot initially impacted the participants' ratings of the physical and temporal dimensions. In week 1, while the robot-assisted task was considered less physically strenuous, the participants found the time taken to be burdensome. The trend for the physical dimension continued in the subsequent weeks. However, the impact of the robot on the temporal dimension diminished, suggesting that participants found it easier to use the robot to complete the tasks in weeks 5 and 8. Furthermore, participants found that the use of robot required less effort in the last week, suggesting that there was a learning effect, and that participants were able to use the robot more efficiently as time progressed. This was also seen in the manner that the participants reported they used the robot and as well as in their observed usage. In week 2, participants would load themselves and the robot and then follow the robot to the living to unload it. They would then return to the kitchen

with the robot. In subsequent weeks, participants would be more likely to not wait for the robot, but rather move around the robot and only load/unload it if they happened to be in the same space as it. This approach employed the robot more efficiently as a supplement to their own capabilities. For the cognitive prosthetic task, the impact of the robot was less clear-cut. Participants rated doing the task with the robot as requiring less mental workload, and this effect persisted throughout the trials. In addition, participants felt that doing the task with the robot required less effort. Despite this, participants rated the use of the robot as requiring more workload in order to perform the task successfully. There was a trend suggesting that for weeks 2 and 5 the use of the robot was seen as more time-consuming; it was also seen as more frustrating across all the trials.

This suggests that despite the experienced benefit of using the robot in this task, there were still associated problems that made it more time-consuming and frustrating.

Research Question 3—The Experience of the Task

The descriptive analysis of the open-ended questions allowed for a deeper and more thorough perspective about the tasks and how they were experienced by the users.

When discussing the initial tasks, participants referenced the constraints imposed on them. Many of their suggestions for making the task easier involved the removal of these constraints. In the cognitive prosthetic task, the participants also considered the means through which they could access the information on the robot as one of the constraints.

In addition, the results from the TLX along this task were mirrored in the way that participants reasoned about the task. Participants described the robot as slow and inflexible and expressed a need to change the way that the robot was used in the task, either by changing how information was presented or by changing the usage of the robot. This was a reflection not just on their experience of the robot, but also how their increased mastery of the task made them consider the role of the robot differently. This even led to two of the participants using the robot in a manner unintended by the experimenters. They asserted control by subverting its use and using the Fetch and Carry functionalities to aid in the Cognitive Prosthetic task. For the remainder of the sample, however, there seemed to be a tacit understanding of a trade-off between the lack of human error in this task and the lack of control. In the Fetch and Carry task, however, despite similar descriptors of the robot being used in terms of it slowing down the task, participants adapted their use of the robot. This allowed the participants to

work around these shortcomings and receive beneficial assistance from the robot. The changes that the participants mainly wanted to implement in terms of how they interacted with the robot were mainly quantitative changes: giving it more space to carry things and letting it move more quickly, in contrast to the changes in the quality of assistance that were suggested in the Cognitive Prosthetic task. It also emerged that, unlike in the Cognitive Prosthetic task, users referenced the robot as a partner and companion in the Fetch and Carry task. This may reflect the open-ended nature of this interaction, and the opportunity for a natural synchronization of behavior to occur gradually. Stienstra and Marti (2012) suggest this is a key factor in developing feelings of sociality and empathy in an HRI situation.

Ecological Validity

The narrative framing of the interactions within the Robot House environment enabled participants to evaluate their interactions in a more relevant and applicable manner than what would have been possible in a traditional laboratory study. Despite the fact that the constrained tasks were part of an experimental study where the participants' interaction with the robot was tightly controlled, there are several factors that support the ecological validity of this study. These tasks were based on the needs of the user personas, and expected interactions arising from these needs. The parallels between observed behaviors and similar interactions with technologies in everyday settings were also encouraging. In the Fetch and Carry task, the process the participants went through when completing the tasks with the robot was quite similar to that of the user in the Hüttenrauch and Eklundh (2002) study. In both cases, the users started off by coordinating their behavior closely with the robot, for example, walking with the robot, and synchronizing their own behavior with that of the robot. They then progressed to using the robot in a more asynchronous manner, with less constant control of the robot. These similarities in interactional outcomes support the notion that many of the qualities of a real-world usage of a final stage prototype were successfully translated into the experimental setup. For the Cognitive Prosthetic task, the manner in which the participants described the role of the robot in the task had elements that map well onto how people perceive such technologies in real-world settings. The issues of autonomy and control come up in both theoretical and practical discussions of the use of robotic technologies (Anderson and Anderson 2008; Sharkey and Sharkey 2012). In particular, the resolution of control issues by subverting assistive technologies has also been reported in real-world settings (Loe 2010), and an analogous

process took place within the experiments. This suggests that for the cognitive prosthetic tasks, many of the salient aspects of using such technologies could be effectively conveyed through this constrained method.

Implications

The findings highlight the need for a user-centered approach to assistive technologies intended for domestic use. The results from the constrained task experiments strongly stress the need for such assistance to allow for personalization and for the robot assistance to be gradually scaled in order to account for changes in task mastery in the user and for coping strategies that the user may adopt. The TLX scores for the Cognitive Prosthetic tasks suggest that total experienced workload may increase where such scaling and alteration of assistance do not occur, due to frustration and disruption to learned coping strategies, despite the robot's assistance being still considered useful. In addition, the open-ended responses to this task suggested that participants came to regard the robot's assistance as hindering their preferred solution to the task. The scores for the Fetch and Carry task, on the other hand, represent a scenario where the roles of both the robot and the participants were less strongly defined. This left a lot of room for mutual adaptation, which in turn led to a more successful interaction in terms of the TLX scores, and also in terms of the participants' reasoning about the task and the role of the robot. This suggests that even in constrained tasks, such as the ones presented here, there is a hedonic dimension to interactions that has a role equal to their purely task- and workload-related aspects. This hedonic quality may be impacted by anthropomorphic interaction capabilities, and an interesting future strand of research into task-related domestic human-robot interaction would be to investigate the role of such capabilities in how users respond to performing tasks with robots.

CONCLUSIONS

The work presented in this article has shown the validity of interaction prototyping, in terms of both a high-level narrative approach in which the participants is involved in the playing of the role of a user of a more "mature" version of the technology being prototyped, and that of separating the task-aspect of such interaction. This two-pronged approach to interactions with future and emerging technologies for the purposes of early prototyping is a valid tool for gaining insight into how such interactions may be experienced by the intended users. The findings in this study have allowed us to replicate findings of real-world studies in terms of how participants reason about their potential adoption of such technologies, as well as to quantify the impact of

assistance in such tasks using the NASA TLX and to highlight issues relevant for the UH Robot House Scenario and human-robot interaction in general.

The work described in this article showed how to successfully integrate constrained tasks (as part of controlled experiments) with more "natural," open-ended scenarios as part of a long-term study into home companion robots operating in domestic environments. We pointed out experimental and methodological challenges and how they have been addressed in this study. The constrained tasks were based on commercially available tools and as such could potentially be used and replicated by other researchers. Being able to share, replicate, and build upon each others' results remains one of the big challenges in human-robot interaction, which otherwise remains in danger of staying a widely fragmented field with different research groups using different robotic platforms, scenarios, and methodological approaches (Dautenhahn 2007). We therefore hope that, in addition to presenting concrete results from a long-term human-robot interaction study, this article has also raised awareness of the main challenges as well as opportunities in the design of interaction technology that supports long-term human-robot interaction.

NOTE

1. Previous studies examining the application of biologically inspired expressive behaviors to Sunflower had shown that participants found the robot's non-anthropomorphic communicative behavior very effective in terms of conveying the robot's intention (Koay et al., 2013).

ACKNOWLEDGEMENTS

We would like to thank our colleagues Michael L. Walters and Joe Saunders for helpful comments, and Fotios Papadopoulos for his help with the AIBO robot.

FUNDING

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement 287624, the ACCOMPANY project, and grant agreement 215554, the LIREC (LIving with Robots and intERactive Companions) project.

REFERENCES

- Anderson, M., and S. L. Anderson. 2008. Ethical healthcare agents. In *Advanced computational intelligence paradigms in healthcare-3*, ed. M. Sordo, S. Vaidya, L. C. Jain, 233-57. Berlin, Germany: Springer.

- Bartneck, C., C. Rosalia, R. Menges, and I. Deckers. 2005. Robot abuse—A limitation of the media equation. *Proceedings of the Interact 2005 Workshop on Agent Abuse*, September, Rome, Italy, 54–7.
- Bickmore, T., and J. Cassell. 2005. Social dialogue with embodied conversational agents. In *Advances in natural multimodal dialogue systems*, ed. J. C. J. van Kuppevelt, L. Dybkjær, and N. Ole Bernsen, 23–54. Berlin, Germany: Springer.
- Bickmore, T. W., L. Caruso, K. Clough-Gorr, and T. Heeren. 2005. 'It's just like you talk to a friend': Relational agents for older adults. *Interacting with Computers* 17(6): 711–35. <http://dx.doi.org/10.1016/j.intcom.2005.09.002>
- Blythe, M. A., and P. C. Wright. 2006. Pastiche scenarios: Fiction as a resource for user centred design. *Interacting with Computers* 18(5): 1139–64. <http://dx.doi.org/10.1016/j.intcom.2006.02.001>
- Buchenau, M., and J. F. Suri. 2000. Experience prototyping. In *Proceedings of the 3rd conference on designing interactive systems: Processes, practices, methods, and techniques*, 424–33. New York, NY: ACM.
- Carroll, J. M. 2000. Five reasons for scenario-based design. *Interacting with Computers* 13(1): 43–60. [http://dx.doi.org/10.1016/S0953-5438\(00\)00023-0](http://dx.doi.org/10.1016/S0953-5438(00)00023-0)
- Chang, Y.-N., Y.-K. Lim, and E. Stolterman. 2008. Personas: From theory to practices. In *Proceedings of the 5th Nordic conference on human-computer interaction: Building bridges*, 439–42. New York, NY: ACM.
- Chatley, A. R., K. Dautenhahn, M. L. Walters, D. S. Syrdal, and B. Christianson. 2010. Theatre as a discussion tool in human-robot interaction experiments—a pilot study. In *Advances in computer-human interactions, 2010*, 73–8. New York, NY: IEEE.
- Coradeschi, S., A. Cesta, G. Cortellessa, L. Coraci, J. Gonzalez, L. Karlsson, F. Furfari, A. Loutfi, A. Orlandini, F. Palumbo, et al. 2013. Giraffplus: Combining social interaction and long term monitoring for promoting independent living. In *Human system interaction (HSI), 2013*, 578–85. New York, NY: IEEE.
- Danish Technological Institute. 2012. James—Robot butler. <http://robot.dti.dk/projects/james-robot-butler.aspx> (accessed December 14, 2012).
- Dautenhahn, K. 2007. Methodology and themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems* 4(1): 103–8.
- Dautenhahn, K., M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, A. Sisbot, R. Alami, and T. Siméon. 2006. How may I serve you?: A robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction*, 172–9. New York, NY: ACM.
- Dindler, C., and O. S. Iversen. 2007. Fictional inquiry—Design collaboration in a shared narrative space. *CoDesign* 3(4): 213–234. <http://dx.doi.org/10.1080/15710880701500187>
- Duffy, B. R., G. M. O'Hare, A. N. Martin, J. F. Bradley, and B. Schon. 2003. Agent chameleons: Agent minds and bodies. In *16th international conference on computer animation and social agents, 2003*, 118–25. New York, NY: IEEE.
- Duque, I., K. Dautenhahn, K. L. Koay, I. Willcock, and B. Christianson. 2013. Knowledge-driven user activity recognition for a smart house: Development and validation of a generic and low-cost, resource-efficient system. In *ACHI 2013, The sixth international conference on advances in computer-human interactions*, Nice, France, February 24–March 1, 141–6.
- Fernaes, Y., M. Hakansson, M. Jacobsson, and S. Ljungblad. 2010. How do you play with a robotic toy animal?: A long-term study of Pleo. In *Proceedings of the 9th international conference on interaction design and children*, 39–48. New York, NY: ACM.
- Fussell, S. R., S. Kiesler, L. D. Setlock, and V. Yew. 2008. How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE international conference on human-robot interaction*, 145–52. New York, NY: ACM.
- Hart, S. G. 2006. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50: 904–8. New York, NY: Sage.
- Hart, S. G., and L. E. Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload* 1(3): 139–83. [http://dx.doi.org/10.1016/S0166-4115\(08\)62386-9](http://dx.doi.org/10.1016/S0166-4115(08)62386-9)
- Horne, R., J. Weinman, N. Barber, R. Elliott, M. Morgan, A. Cribb, et al. 2005. *Concordance, adherence and compliance in medicine taking*. London, UK: National Co-ordinating Centre for NHS Delivery and Organization R & D (NCCSDO).
- Huijnen, C., A. Badii, H. van den Heuvel, P. Caleb-Solly, and D. Thiemert. 2011. 'Maybe it becomes a buddy, but do not call it a robot'—Seamless cooperation between companion robotics and smart homes. In *Ambient intelligence*, 324–29. Berlin, Germany: Springer.
- Hüttenrauch, H., and K. S. Eklundh. 2002. Fetch-and-carry with CERO: Observations from a long-term user study with a service robot. In *11th IEEE international workshop on robot and human interactive communication, 2002, Proceedings*, ed. D. Keyson, M. L. Maher, N. Streitz, A. D. Cheok, J. C. Augusto, R. Wichert, G. Englebienne, H. Aghajan, and B. Kröse, 158–63. New York, NY: IEEE.
- Kamsma, Y. P., W. H. Brouwer, and J. P. Lakke. 1995. Training of compensational strategies for impaired gross motor skills in Parkinson's disease. *Physiotherapy Theory and Practice* 11(4): 209–29. <http://dx.doi.org/10.3109/09593989509036407>
- Kidd, C. D., and C. Breazeal. 2008. Robots at home: Understanding long-term human-robot interaction. In *Intelligent robots and systems, 2008, IROS 2008*, 3230–5. New York, NY: IEEE.
- Koay, K. L., G. Lakatos, D. Syrdal, M. Gácsi, B. Bereczky, K. Dautenhahn, A. Miklósi, and M. L. Walters. 2013. Hey! There is someone at your door. A hearing robot using visual communication signals of hearing dogs to communicate intent. In *IEEE Symposium on Artificial Life (ALIFE), 2013*, 90–97. New York, NY: IEEE.
- Koay, K. L., E. A. Sisbot, D. S. Syrdal, M. L. Walters, K. Dautenhahn, and R. Alami. 2007. Exploratory study of a robot approaching a person in the context of handing over an object. In *AAAI spring symposium: Multidisciplinary collaboration for socially assistive robotics*, 18–24. Menlo Park, CA: AAAI Press.
- Koay, K. L., D. S. Syrdal, M. Ashgari-Oskoei, M. L. Walters, and K. Dautenhahn. 2014. Social roles and baseline proxemic preferences for a domestic service robot. *International Journal of Social Robotics* 6(4): 469–88. <http://dx.doi.org/10.1007/s12369-014-0232-4>
- Koay, K. L., D. S. Syrdal, K. Dautenhahn, K. Arent, L. Malek, and B. Kreczmer. 2011. Companion migration—initial participants'

- feedback from a video-based prototyping study. In *Mixed reality and human–robot interaction*, ed. X. Wang, 133–51. Berlin, Germany: Springer.
- Koay, K. L., D. S. Syrdal, M. L. Walters, and K. Dautenhahn. 2009. Five weeks in the robot house—Exploratory human–robot interaction trials in a domestic setting. In *Advances in computer–human interactions, 2009, ACHI'09*, 219–26. New York, NY: IEEE.
- Lammer, L., A. Huber, A. Weiss, and M. Vincze. 2014. Mutual care: How older adults react when they should help their care robot. In *AISB2014: Proceedings of the 3rd international symposium on new frontiers in human–robot interaction*, University of London, April 1–4. London, UK: Goldsmith's.
- Latane, B., K. Williams, and S. Harkins. 1979. Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology* 37(6): 822–32. <http://dx.doi.org/10.1037/0022-3514.37.6.822>
- Loe, M. 2010. Doing it my way: Old women, technology and wellbeing. *Sociology of Health & Illness* 32(2): 319–34. <http://dx.doi.org/10.1111/j.1467-9566.2009.01220.x>
- Modayil, J., R. Levinson, C. Harman, D. Halper, and H. Kautz. 2008. Integrating sensing and cueing for more effective activity reminders. In *AAAI Fall 2008 Symposium on AI in Eldercare: New solutions to old problems*, 7–9. Menlo Park, CA: AAAI Press.
- Newell, A. F., A. Carmichael, M. Morgan, and A. Dickinson. 2006. The use of theatre in requirements gathering and usability studies. *Interacting with Computers* 18(5): 996–1011. <http://dx.doi.org/10.1016/j.intcom.2006.05.003>
- Parlitz, C., M. Hägele, P. Klein, J. Seifert, and K. Dautenhahn. 2008. Care-O-Bot 3—Rationale for human–robot interaction design. *Proceedings of 39th International Symposium on Robotics (ISR)*, Seoul, Korea, October 15–17, 275–80.
- Payr, S. 2010. Closing and closure in human–companion interactions: Analyzing video data from a field study. In *2010 IEEE RO-MAN*, 476–81. New York, NY: IEEE.
- Rubio, S., E. Díaz, J. Martín, and J. M. Puente. 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology* 53(1): 61–86. <http://dx.doi.org/10.1111/j.1464-0597.2004.00161.x>
- Russo, J. E., E. J. Johnson, and D. L. Stephens. 1989. The validity of verbal protocols. *Memory & Cognition* 17(6): 759–69. <http://dx.doi.org/10.3758/BF03202637>
- Schwartz, D., M. Wang, L. Zeitz, and M. E. Goss. 1962. Medication errors made by elderly, chronically ill patients. *American Journal of Public Health and the Nation's Health* 52(12): 2018–2029 <http://dx.doi.org/10.2105/AJPH.52.12.2018>
- Seland, G. 2009. Empowering end users in design of mobile technology using role play as a method: Reflections on the role-play conduction. In *Human centered design*, ed. M. Kurosu, 912–21. Berlin, Germany: Springer.
- Sharkey, A., and N. Sharkey. 2012. Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology* 14(1): 27–40. <http://dx.doi.org/10.1007/s10676-010-9234-6>
- Stienstra, J., and P. Marti. 2012. Squeeze me: Gently please. In *Proceedings of the 7th Nordic conference on human–computer interaction: Making sense through design*, 746–50. New York, NY: ACM.
- Sung, J.-Y., R. E. Grinter, H. I. Christensen, and L. Guo. 2008. Housewives or technophiles?: Understanding domestic robot owners. In *2008 3rd ACM/IEEE International Conference on human–robot interaction (HRI)*, 129–36. New York, NY: IEEE.
- Syrdal, D. S., K. Dautenhahn, K. L. Koay, and W. C. Ho. 2014. Views from within a narrative: Evaluating long-term human–robot interaction in a naturalistic environment using open-ended scenarios. *Cognitive Computation* 6(4): 741–59.
- Syrdal, D. S., K. Dautenhahn, K. L. Koay, and M. L. Walters. 2009. The negative attitudes towards robots scale and reactions to robot behaviour in a live human–robot interaction study. *New frontiers in human–robot interaction, a symposium at the AISB2009 Convention*, Edinburgh, UK, April 6–9.
- Syrdal, D. S., K. Dautenhahn, M. L. Walters, K. L. Koay, and N. R. Otero. 2011. The theatre methodology for facilitating discussion in human–robot interaction on information disclosure in a home environment. In *2011 IEEE RO-MAN*, 479–84. New York, NY: IEEE.
- Syrdal, D. S., K. L. Koay, M. Gácsi, M. L. Walters, and K. Dautenhahn. 2010. Video prototyping of dog-inspired non-verbal affective communication for an appearance constrained robot. In *2010 IEEE RO-MAN*, 632–7. New York, NY: IEEE.
- Syrdal, D. S., K. L. Koay, M. L. Walters, and K. Dautenhahn. 2009. The boy-robot should bark!—Children's impressions of agent migration into diverse embodiments. In *Proceedings of the new frontiers in human–robot interaction, a symposium at the AISB2009 Convention*, Edinburgh, UK, April 6–9.
- Syrdal, D. S., N. Otero, and K. Dautenhahn. 2008. Video prototyping in human–robot interaction: Results from a qualitative study. In *Proceedings of the 15th European conference on cognitive ergonomics: The ergonomics of cool interaction*, 29–35. New York, NY: ACM.
- Tapus, A., C. Tapus, and M. J. Matarić. 2008. User–robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics* 1(2): 169–83. <http://dx.doi.org/10.1007/s11370-008-0017-4>
- Walker, J. E., and J. Howland. 1991. Falls and fear of falling among elderly persons living in the community: Occupational therapy interventions. *American Journal of Occupational Therapy* 45(2): 119–22. <http://dx.doi.org/10.5014/ajot.45.2.119>
- Walters, M. L., M. Lohse, M. Hanheide, B. Wrede, D. S. Syrdal, K. L. Koay, A. Green, H. Hüttenrauch, K. Dautenhahn, G. Sagerer, and K. Severinson Eklund. 2011. Evaluating the robot personality and verbal behavior of domestic robots using video-based studies. *Advanced Robotics* 25(18): 2233–54. <http://dx.doi.org/10.1163/016918611X603800>
- Walters, M. L., M. A. Oskoei, D. S. Syrdal, and K. Dautenhahn. 2011. A long-term human–robot proxemic study. In *2011 IEEE RO-MAN*, 137–42. New York, NY: IEEE.
- Yagoda, R. E. 2010. Development of the human robot interaction workload measurement tool (HRI-WM). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54: 304–8. New York, NY: Sage.