"Was that successful?" On Integrating Proactive Meta-Dialogue in a DIY-Assistant using Multimodal Cues

Matthias Kraus
Ulm University
matthias.kraus@uni-ulm.de

Marvin Schiller
Robert Bosch GmbH
marvin.schiller@de.bosch.com

Gregor Behnke*
University of Freiburg
behnkeg@cs.uni-freiburg.de

Pascal Bercher*
The Australian National University
pascal.bercher@anu.edu.au

Michael Dorna
Robert Bosch GmbH
michael.dorna@de.bosch.com

Michael Dambier
Robert Bosch GmbH
michael.dambier@de.bosch.com

Birte Glimm
Ulm University
birte.glimm@uni-ulm.de

Susanne Biundo
Ulm University
susanne.biundo@uni-ulm.de

Wolfgang Minker
Ulm University
wolfgang.minker@uni-ulm.de

ABSTRACT

Effectively supporting novices during performance of complex tasks, e.g. do-it-yourself (DIY) projects, requires intelligent assistants to be more than mere instructors. In order to be accepted as a competent and trustworthy cooperation partner, they need to be able to actively participate in the project and engage in helpful conversations with users when assistance is necessary. Therefore, a new proactive version of the DIY-assistant Robert is presented in this paper. It extends the previous prototype by including the capability to initiate reflective meta-dialogues using multimodal cues. Two different strategies for reflective dialogue are implemented: A progress-based strategy initiates a reflective dialogue about previous experience with the assistance for encouraging the self-appraisal of the user. An activity-based strategy is applied for providing timely, task-dependent support. Therefore, user activities with a connected drill driver are tracked that trigger dialogues in order to reflect on the current task and to prevent task failure. An experimental study comparing the proactive assistant against the baseline version shows that proactive meta-dialogue is able to build user trust significantly better than a solely reactive system. Besides, the results provide interesting insights for the development of proactive dialogue assistants.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **HCI design and evaluation methods**; **User studies**; *Interaction design theory, concepts and paradigms.*

*The work described in this paper was developed while still being at Ulm University.

ICMI '20, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-7581-8/20/10…$15.00
https://doi.org/10.1145/3382507.3418818

KEYWORDS

assistance system; proactivity; meta-dialogue; do-it-yourself

ACM Reference Format:
Matthias Kraus, Marvin Schiller, Gregor Behnke, Pascal Bercher, Michael Dorna, Michael Dambier, Birte Glimm, Susanne Biundo, and Wolfgang Minker. 2020. "Was that successful?" On Integrating Proactive Meta-Dialogue in a DIY-Assistant using Multimodal Cues. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20), October 25–29, 2020, Virtual event, Netherlands.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3382507.3418818

1 INTRODUCTION

Do-it-yourself (DIY) home improvement is one of the most popular hobbies around the world. A study conducted in 2019 revealed that, in Germany alone, 11.91 million people aged 14 and older showed particular interest in the topic of home improvement [1]. However, for novices the entrance into the DIY-domain is, despite interest, fraught with difficulties. For example, beginners do not have specific knowledge about materials and tools. Thus, they lack the technical and practical abilities for conducting DIY projects without proper instructions. Furthermore, due to being afraid of failure and damage, they often have low self-confidence and shy away from using typical tools such as an electric drill and saw. Therefore, novices are in need for trusted and competent assistance in order to receive appropriate guidance and tutoring. One possibility to obtain such assistance are interactive assistant systems, e.g. intelligent tutoring systems (ITS) [2, 22] or companion systems [16]. In previous work, a conversational assistant (CA) named Robert has been developed in collaboration with Robert Bosch GmbH, one of the leading manufacturers of power tools [12]. Robert provides novice DIYers with planner-generated step-by-step instructions on how to successfully complete a given DIY task. For illustration of individual task steps, instructional texts, images, and videos are automatically selected using knowledge-based reasoning. In order to have dialogue capabilities, the assistant is able to process multimodal natural language (text, speech). A study comparing the first version of Robert with a non-interactive slide-show-based baseline showed that novice users were favourable towards the assistance of Robert. A drawback of the system was that it could only provide reactive assistance, i.e. on user request. Therefore, it could not support the user on its own initiative in situations where system interventions could have been beneficial. Hence, the assistant lacked the human like capability of proactivity. Current research of proactive behaviour in human-computer interaction (HCI) suggests that active system actions, if provided appropriately, have a positive effect on the user's acceptance and trust towards the system [3, 40, 59, 63]. For this reason, they should be considered in the design process of a CA. However, finding the right strategies for situation-dependent proactivity is far from being solved [55].

Hence, this paper makes the following contributions: First, a new iteration of ROBERT is equipped with situation-adaptive proactive functionality. Therefore, strategies for meta-dialogue initiation are implemented. The term meta-dialogue is used for task-independent interaction that relates, for example, to correction of user actions or asking for feedback about the system's behaviour [69]. A *progress-based* strategy initiates a reflective dialogue about previous experience with the assistance after completion of important sub-tasks. Reflective dialogue has shown to be useful in order to boost the development of conceptual knowledge, problem-solving ability, and overall experience with intelligent tutoring systems [33, 34, 60]. Additionally, an *activity-based* strategy is applied for providing timely, task-dependent support. Therefore, the user's activities with an electric drill driver are tracked. By applying additional sensors and machine-learning, it is possible to classify seven different activities, which then are used for dialogue initiation in order to reflect on the current task and to prevent task failure. Tool usage, as an additional input modality, can be seen as a multimodal cue for engaging in a multimodal dialogue in combination with typical user input (speech, text, touch). Furthermore, a between-subject experimental study is presented, comparing the proactive variant of ROBERT with the reactive variant in terms of acceptance, perceived trustworthiness, and usability. During the study, subjects are instructed to build a wooden key rack in cooperation with the assistant. The results show that proactive meta-dialogue is able to build user trust significantly better than a solely reactive system. Besides, the results provide interesting insights for the development of proactive dialogue assistants, especially with regard to the user's gender.

The outline of this paper is as follows: Related work regarding CA systems, proactive HCI, and human-computer trust is presented in Section 2. Section 3 provides an overview of the system's components, while Section 4 deals with the integration of proactive meta-dialogue in ROBERT. Subsequently, the experimental setup and evaluation methods are described in detail. In Section 6, the results of the study are presented. A discussion of found results is provided in Section 7. Finally, the paper is concluded in Section 8.

2 RELATED WORK

Systems that mimic human conversation using text or spoken language are known as conversational agents, such as Apple Siri and Amazon Alexa [48]. CAs extend the abilities of agents by additionally providing guidance and help with problem-solving, decision-making, and learning [41]. Successful applications of CAs are numerous and span diverse fields like health care [41], education [36], or assistance in the public (museum guiding [37]) and private sector (cooking assistant [52]). A sophisticated realisation of CAs can be found in the line of research of companion-technology for technical cognitive systems [16]. According to the authors' definition, companion systems combine the cognitive abilities of planning, knowledge reasoning, and dialogue in order to provide individual and adaptive assistance. Empirical evaluations of companion systems [12, 15] demonstrate the acceptance and usefulness for users in real-world application scenarios. From a dialogue research point of view, the alignment of artificial intelligence (AI) and human problem-solving, i.e. the integration of AI planning in user-centred dialogue, has been studied in the context of companion systems [53]. Furthermore, insights about the impact of explanations of system behaviour in incomprehensible situations during interaction with companions and their effect on trust have been provided [54]. These explanations mainly serve a social (interactional) purpose [65] in building positive social relations by building trust and rapport through transparency [17, 19]. In the context of companion systems, social interaction is necessary as the user has to form a bond with the system in order to accept and trust its instructions. Therefore, the repertoire of social conversation is extended with reflective dialogues in this paper. The intent of reflective dialogue is for the user to understand what has occurred and to actively inquire about his actions in order to process impactful experiences and to learn [21]. Reflection is an important part of action research, among planning, action and observation [47], and has shown to have a positive impact on learning [34]. Most previous research on reflective dialogue has been contributed in the field of intelligent tutoring systems [33–35, 60]. Katz et al. [35], for example, present a comparative analysis of problem-solving and reflective dialogue. By studying the effect of post-practice reflection (PPR), the authors show that student self-explanation and self-appraisal are more prevalent in PPR and correlate strongly with learning. Furthermore, they show that a debrief after task completion correlated with an effective performance.

In this paper, the concept of reflective dialogue is investigated in the context of companion systems. However, the focus is shifted away from the impact on learning towards the impact on the social human-companion relationship. Here, the human-computer trust relationship is considered. Additionally, it is deemed inevitable for the reflective dialogue to be initiated by the assistant. For this reason, an overview of proactive HCI is presented in the next section.

The term proactivity stems from the domain of occupational and organisational psychology [23, 58]. Although there are a lot of ways to describe proactive behaviour, they all have in common that proactive behaviour is about anticipatory action in order to influence future situations and to prevent problems. Nothdurft et al. [55] thereupon derived a definition for dialogue systems: "Proactivity in technical systems is an autonomous, anticipatory system-initiated behaviour, with the purpose to act in advance of a future situation, rather than only reacting to it". Additionally, the authors stated three challenges for proactive dialogue systems: If proactivity is necessary, when, and how proactivity should be initiated. The necessity of proactive behavior depends on several factors. Most notably, the decision should be utility-based and help the user in general, rather than aim solely at task success [51]. Therefore, reflective meta-dialogue is implemented in ROBERT as a proactive dialogue strategy. In doing so, an enhanced trust-building towards the user is expected. However, this is only beneficial for the interaction, if the style and timing of proactive behaviour is designed appropriately. Initiating an interaction at an inappropriate point of time and in the wrong way could be perceived as disruptive and obtrusive. This may lead to a corrupted human-computer relationship, especially regarding the user's perceived trust in the system [30, 46].

When a system should initiate proactive behaviour is closely related to its ability to anticipate, i.e. to make assumptions about the future. For assistant systems, there exist several possibilities to predict future situations, e.g. gaze of the user [28], or verbal cues [66]. In case of CAs, making assumptions about the user and the situation requires extensive knowledge about the user's current activity state. Activity recognition is an emerging research field. Especially, sensor-based approaches have made great progress due to the advancements in deep learning [72]. Sensors for activity recognition can be divided roughly into the categories body worn sensors (smartphone, wearable), object sensors (RFID, accelerometer), ambient sensors (sound, door sensor), and hybrid variants. They are applied in various application fields, for example, daily life monitoring [68], medical environments [18, 24], and smart homes [43, 61]. In this work, object sensors (gyroscope, accelerometer, compass) integrated into an electric screwdriver of the Robert Bosch company are used to infer the user's current activity. The activity state is used for proactive dialogue initiation in order to reflect on the current task and to prevent task failure. Furthermore, proactive behaviour is triggered accordingly after certain points of progress during the project. The way in which a proactive system should initiate its actions is a current hot topic in HCI and mostly related to the intervention style (e.g. see Wagner et al. [71]) or a system's level of autonomy [3, 59, 73]. In related literature, the system's autonomy is frequently categorised into 10 levels [4, 63, 67]. These levels can be roughly abstracted to low- (computer offers no assistance), medium- (computer offers a suggestion), and high-level (computer executes a suggestion) proactivity. Recent studies have shown that a medium-level of proactivity can be beneficial to the human-computer trust relationship [40, 63]. Therefore, proactive behaviour of Robert is implemented to be suggestion-based, where the user is still in control of the interaction.

The concept of trust is fundamental both in interpersonal as well as human-computer relationships [42, 45]. Trust can be defined as "the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability" [42]. Substantial research on this topic has been provided in the field of autonomous systems [26, 50, 57] and human-robot interaction [20, 25]. In this paper, the trust model developed by Madsen and Gregor [44] serves as the foundation for the evaluation of Robert's proactive behaviour. This hierarchical model is built on five basic components of trust: Personal attachment and faith form the bases for affect-based trust while perceived understandability, perceived technical competence, and perceived reliability constitute the bases for cognition-based trust. Affect-based trust refers to a long-term human-computer relationship, being established through frequent interactions with a system. Conversely, cognition-based trust refers to a more short-termed trust. Here, mostly the functionality and usability of a system are of importance. The trust model by Madsen and Gregor [44] allows for a differentiated measurement of which factors of trust are particularly influenced by the manipulation of Robert's proactive behaviour.

3 SYSTEM OVERVIEW

Robert comprises three components for providing suitable assistance to novice DIYers [12, 13, 39]: a planning-, an ontology-, and a dialogue-management-component. An overview of the workflow between these components is depicted in Fig. 1. All three components share the same model information. However, each component only stores the information for handling which it is best suited for. When required, information is transmitted from one component to another. To allow for this interoperability and to ensure coherent storage of models and information, a specific modelling paradigm is used, which e.g. allows for storing parts of the planning model in a structured way in the ontology [64].

3.1 HTN Planner

Robert proposes a course of action to its user that if performed will complete the DIY project the user wants to undertake. Robert's planning component is responsible for determining this course of action – called a *plan*. For determining a suitable plan, the planner utilises a general description of the DIY setting and the user's project in terms of a *planning model*. This model encompasses formal descriptions of the available tools and materials as well as the actions that can be used to manipulate the environment in a DIY setting – e.g. sawing, drilling, fixating. The model itself does not pertain to the specifics of a single (or some) specific problems or projects, but instead is a general description of the possible activities that can be performed in a DIY setting. This generality allows Robert's planner to flexibly adapt its plan to the current situation and project of the user. For example, it can come up with other means of making a large hole, if no Forstner bit is available. For formalising the model, we use the concept Hierarchical Task Network (HTN) planning [14]. Using HTNs enables us to suitably integrate the planning model with the ontology [11, 64] s.t. information is stored only once and handled by the component best suited for it. Further, the integration of the HTN planning model and the ontology enables us to answer questions that connect the knowledge from both models [5]. Lastly, the hierarchical nature of the description allows Robert to provide abstract instructions in addition to detailed instructions. This is useful if a user is already familiar with some procedures in the DIY setting (e.g. pre-drilling) – and thus does not need to be instructed on how to perform them again. Robert uses a SAT-based planner to find optimal (shortest) plans [7–10].

3.2 Ontology Manager

The ontology manager organises Robert's static knowledge specific to the DIY domain. DIY tools and objects (e.g. drills, bits, saw blades, . . .) are organised in an ontology and characterised by properties such as colour, shape, but also technical parameters (e.g. battery voltage) and functionalities (e.g. that a drill driver can serve both as a drill and as a screw driver). The ontology also stores suitable configurations for instantiating actions in the domain, e.g. the recommended speed settings for drilling in wood. The ontology's DIY domain model is provided both to the planner and the dialogue manager. In addition, the ontology manager organises the instruction elements (texts, images, videos) from which the step-by-step instructions are to be assembled according to the actions and

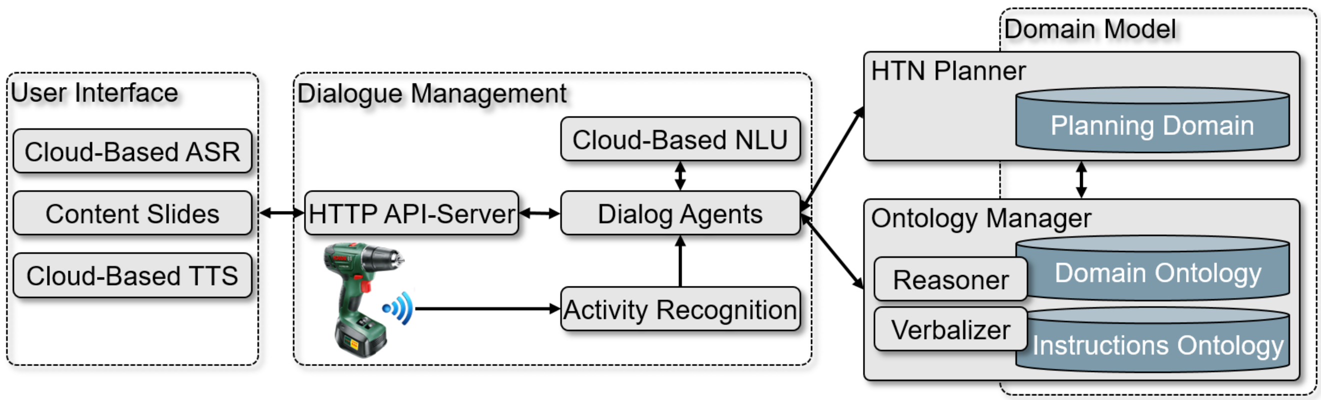

Figure 1: Overview of ROBERT's architecture. A user interacts with the assistant's interface capable of multimodal input recognition. User input is forwarded to a server-based dialogue manager that mediates the interaction with the HTN planner and the ontology manager. In addition, the system is able to track the user's activity with a connected electric drill for proactive dialogue initiation (product picture: Bosch).

parameters instantiated by planning, for which logical reasoning (classification) in the ontology is used. The ontology manager is also queried when ROBERT answers factual questions from the user, for which text is generated from the stored descriptions and media are retrieved (as described in the following section, see also [12]).

3.3 Dialogue Manager

ROBERT's dialogue manager controls the interaction between the assistant and its user. It is implemented as a web-based server application following a client-server architectural approach. On the client side, a user interacts with a browser-interface based on JavaScript. It presents the generated plan in the form of step-by-step instructions (see Fig. 2). The content of the instructions is provided in the form of a textual and visual (picture, video-on-demand) task description. The interface is capable of processing multi-modal user input (speech, touch, text). Spoken language is transformed into text using Google Chrome's web speech API. Each new input is forwarded to the dialogue-server using HTTP-methods (POST, GET, PUT) and JSON as data format.

For dialogue management, an agent-based approach is utilised [62]. For each component of ROBERT there exists an individual dialogue agent that is handling module-specific tasks. In order to recognise the user's intention, i.e. which component is best suited for processing the user's request, the statistical-driven natural language understanding (NLU) unit *Rasa* (rasa.com) is applied. For example, when the user has a plan-related intention, such as getting instructions for a specified DIY project or the wish to modify the plan according to his preferences, the HTN planner is invoked by its respective dialogue agent. The planner then generates a sequence of actions, i.e. the plan, providing appropriate instructions. The ontology is used to augment the symbolic description of the actions with textual instructions and media contents for presentation to the user. The ontology-related agent is able to handle requests for information and explanation about specific materials and tools, e.g. *"What is a drill-bit?"* or *"What does a drill-bit look like?"* The returned explanation is presented to the user in the form

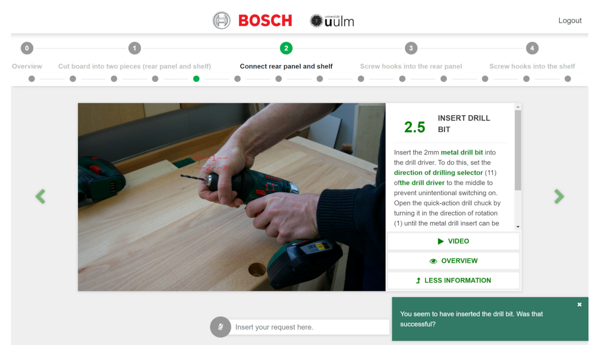

Figure 2: Screenshot of the interface.

of a pop-up modal containing a textual and/or visual description. Additionally, the written descriptions are synthesised to speech using Chrome's TTS API. The dialogue-related agent is responsible for meta-dialogues (information confirmation/grounding, reflection) and for keeping track of the project and dialogue state, stored together in an interaction state *S*. This state *S* contains information about the project step (*Action*) the user is currently working on, as well as past user input (*UserAct*), and system output (*SystemAct*). The system uses this information for adequate response generation.

4 INTEGRATION OF PROACTIVE META-DIALOGUE

Based on the described version of ROBERT two proactive meta-dialogue strategies are developed for providing an additional way of assistance: Following a *progress-based* strategy, proactive system behaviour is triggered at the end of important sub-tasks. These moments mark a clear cut between two main sub-tasks, enabling the user to pause, reflect, and prepare for the next sub-task. For example, "cutting the wooden board in two pieces (rear and shelf)" is the first main task of building the key rack in the evaluation scenario. As a last step of this task, the user is instructed to loosen the screw clamps that are used for fixating the wooden board. There, a reflective meta-dialogue is initiated after the user has spent ten seconds on the instructional slide. This timing is chosen heuristically based on the average reading duration of the instruction. An example dialogue of this interaction scenario between the system (**S**) and its user (**U**) is provided in the following:

S: The first step to building a key rack has been taken. Are you satisfied so far?
U: Yes, everything went fine.
S: What did you like about this craft step?
U: I enjoyed using the electric saw. Never thought I could handle it so easily.
S: How did you like the instruction up to here?
U: Nothing to complain right now.
S: Thanks for your response. Would you like to know the steps for the next sub-task?
U: Yes! Thanks!

As can be seen in the example, each reflective meta-dialogue consists of a maximum of four turns: Questioning the overall satisfaction, asking about an appraisal of the finished craft step and the instruction, and proposing the initiation of the next project step. The interaction follows a static dialogue flow, where the NLU is only applied for confirmations. Here, the system expects affirmations or denials from the user, and uses reprompts otherwise. For the appraisal questions users could provide arbitrary answers. Questioning the overall satisfaction is intended to boost self-awareness about the finished task. In combination with the specific question about the likeability of the completed craft step, this should animate the user to self-appraise, and to appreciate the made progress. As self-appraisal has shown positive results on users of tutoring systems regarding learning and task performance [35], this is also deemed to positively affect the perceived assistance of ROBERT. Questioning the likeability of the provided instructions is intended to set the user's focus on the competence and reliability of the system with regard to its helpful guidance. This is considered to manifest trust in the assistant. Afterwards, the user is proposed to move on to the next task which finishes the reflective meta-dialogue. In summary, the reflective meta-dialogue serves the purpose to get users to talk to the system in order to establish a trust-relationship by letting the system appear to be careful and competent. Additionally, it should reinforce the users' thinking about their capabilities and positively contribute to their experience with the system. It is possible to ignore the ROBERT's proactive behaviour or to quit the dialogue at each step in order to continue with the project. In doing so, the assistant lets users unobtrusively in control of the interaction, which has shown to benefit the human-computer trust relationship in proactive dialogue systems [40, 63]. The wording of the system's utterances is alternated for each sub-task to increase naturalness.

Following an *activity-based* strategy, situation-dependent proactive behaviour is triggered using cues from a connected drill serving as an external sensor of user behaviour [6]. To proactively provide assistance, sensor data for tracking the user's current activity is used, e.g. drilling or screwing. To collect sensor data, an inertial measurement unit is integrated into a standard cordless drill driver and connected to a Wi-Fi development board, such that gyroscopic, accelerometric and compass data are transmitted from the device. Activity classification is provided by a neural network trained with data from 12 participants collected in a separate data-collection experiment. A deep neural network approach is preferred since it is considered state-of-the-art in human activity recognition and yields good results for classifying movement patterns based on sequences of raw sensor inputs (e.g. see Ordóñez and Roggen [56]). Classification serves to distinguish the following classes of (in-)activity: *off* (machine not moved), *screwing*, *drilling*, *drill change*, *battery change*, *in use* (machine is moved, motor is off), and *other*. Activity classification reached an accuracy of 0.91 (in 4-fold cross validation, micro avg. F1 = 0.91, macro avg. F1 = 0.83) in the classifier-training experiment [29]. Additional information is available in the form of probability distributions, e.g. of the current activity, the activity's operation time and frequency of its occurrence.

ROBERT receives new information about the user's activities every 500 ms via a TCP-socket connection. For leveraging this information during the interaction, the concept of a *machineState* as a part of the interaction state S is introduced. The *machineState* allows to trigger proactive system actions based on implicit knowledge about the user's tool activity. It consists of so-called *machineActs*. An act has information about the current activity, its operation time, and how often this act has occurred before. The *machine-State* is updated each time a different user activity is tracked with a high confidence, i.e. the activity's probability has to be higher than 0.95.

Activity-based proactive system behaviour addresses the active initiation of a reflective meta-dialogue with users in order to check whether they are performing the project's steps correctly and to provide help in the case of failure. For this reason, a rule-based decision model for intervention is implemented. At each project step where the connected tool needs to be applied, ROBERT is able to actively start a dialogue with its user. Depending on the context it is differentiated between three different interaction scenarios: A message is triggered after the user picked up the machine (*in use*-classification): *"You seem to be working with the connected tool for the first time. Don't worry, I'll guide you through the steps!"* This message is used for making the user aware of the connected tool, and can be triggered only once. To react to possible user insecurity about the current project step, a help request is sent to the user after three minutes of inactivity (*off*-classification). This kind of request is only executed during steps where an action with the connected tool has to be performed: *"I haven't seen any tool activity by you in three minutes. Do you need help?"* In case the user affirms this question, he or she is invited to watch a video of the current step.

Finally, ROBERT checks whether the user performs the instructed task correctly (see Fig. 2). An interaction is initiated whenever one of the activities *battery change*, *drilling*, *drill change*, or *screwing* is recognised. The user is asked whether the current step is going well. If the user answers with *no*, further help is offered. If help is requested, the system starts an explanatory video of the task. Otherwise, the system apologises for the disturbance.

S: I noticed that you were drilling. Was that successful?
U: No, it wasn't.
S: Ok, do you need additional help?
U: Yes, please.
S: A video of this project step could help. I'm going to play it for you. *(Then a video is played.)*

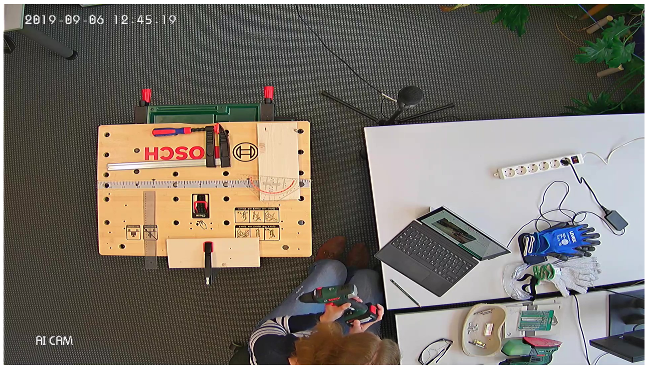

Figure 3: A user attaches the battery of the electric drill driver while being assisted by ROBERT.

5 EXPERIMENTAL SETUP

For comparing the effect of proactive meta-dialogue on the user's perceived trust, acceptance, and experience, a between-subject A/B-test was conducted. Here, the proactivity of ROBERT was used as independent variable, where an assistant capable of proactive meta-dialogue was compared to a baseline variant without any proactive functionalities. Participants were randomly distributed to both study groups.

33 German subjects were recruited by a professional institute for study consulting (MTO). They were evenly assigned to each study group regarding gender and age. A criterion for study participation was to be a novice in DIY projects. However, three subjects had to be excluded from the evaluation, as they were rated as "quite experienced" by study observers in hindsight. In addition, two participants had to be excluded because they did not work according to the study plan. One participant had to be excluded due to malfunction of the system, and one participant aborted the project. This resulted in data from 26 subjects being considered for evaluation. The average age of subjects was 37.54 ($SD = 14.72$). Study participants had various professional and educational backgrounds. The group size of subjects working with the proactive assistant was 12 (5 male, 7 female), while 14 subjects (5 m, 9 f) worked with the baseline version of ROBERT. All subjects received 50 € independently of the study outcome.

After the welcome procedure, the subjects were provided with first instructions and details of the study. They had to read and sign the informed consent and had to fill out a pre-test questionnaire regarding demographics and possible confounding variables, such as previous experience with DIY and technical affinity. Furthermore, propensity towards trust in autonomous systems with the scale by Merritt et al. [49] was measured in order to gain information about the users' initial trust. Afterwards, they received an interactive tutorial in order to get familiar with the standard functionalities of ROBERT, e.g. how to activate the speech recognition, and how to navigate through the user interface. Subsequently, they were asked to build a key rack from a wooden board in cooperation with the DIY-assistant ROBERT. The task consisted of four sub-tasks: sawing a plank into two boards, connecting the boards, attaching two hangers to the back, and adding four hooks to the tray. For the duration of the construction, an experiment facilitator was in the same room as the participant for observation, but was not allowed to assist. The study was captured on video and audio and streamed to a separate room. There, study observers took notes about specific events and participant features, e.g. their subjective assessment of participants' DIY-experience levels. After completion, the participants had to fill in a questionnaire to assess the dependent variables. Total duration of the experiment was between 1.5-2.5 hours. The study setup is depicted in Fig. 3.

In the experiment, the system's perceived trust, acceptance, and the user experience with the system were assessed as dependent variables. For measuring the speech capabilities of the system, the Subjective Assessment of Speech System Interfaces (SASSI) questionnaire [27] was employed. This questionnaire contains the subscales annoyance, user satisfaction, cognitive demand, speed, and habitability. According to the authors a "habitable system may be defined as one in which there is a good match between the user's conceptual model of the system and the actual system." The acceptance towards the assistant was measured with a scale developed by Van Der Laan et al. [70]. To determine trust towards ROBERT, the Trust in Automated Systems Scale [31] was implemented. By measuring the difference between these post-trust ratings and the propensity to trust before the experiment, the effects of the proactive assistant on trust establishment could be observed. Furthermore, scales for measuring the bases of trust developed by Madsen and Gregor [44] were used. All scales were translated into German and slightly modified for study context. All scales were assessed with a 7-point (except where noted with 5-point) Likert scales ranging from 1 = "totally disagree" to 7 = "totally agree" except for the acceptance assessment which used contrary adjective pairs on a 7-point Likert scale.

According to our preceding considerations and related work, the following hypotheses were tested:

H1: Proactive meta-dialogue leads to a better establishment of trust than non-proactive dialogue, i.e. the difference of within-subject measured trust before and after the experiment is significant.

H2: Proactive meta-dialogue leads to higher perceived trust than non-proactive dialogue, i.e. the difference of between-subject measured trust after the the experiment is significant.

H3: Proactive meta-dialogue leads to higher ratings for acceptance, satisfaction, and habitability than non-proactive dialogue.

6 RESULTS

In this paper, results of the effect of proactive meta-dialogue on the user's perception of ROBERT are presented. A previously published paper [6] focused on the evaluation of user-adaptive planning. Both evaluations were examined in the same study setup.

For data analysis, a Mann-Whitney-U-Test was calculated to determine if there were differences in confounding as well as dependent variables. A non-parametric test was chosen, as a rather small sample size was examined. Additionally, a Pearson-Chi-Square test was used for the confounding variables, previous experience in DIY-tools (drilling, sawing,...) and in speech assistants (Siri, Alexa, Google Assistant, ...). These were measured on nominal scales. Confounding group variables for proactive behaviour could be ruled out, as measurements for predisposed trust in autonomous systems [49], technical affinity [32], previous experience in DIY-tools and speech assistants were not significant (all p-values > 0.05). Solely, the Chi-Square test for previous experience with Alexa ($\chi^2 = 7.22,\ p < 0.01$) became significant. However, experience with similar assistants was evenly distributed, hence this result was negligible. In addition, participants' age and gender was similarly distributed for the different experimental groups and no outliers were found in the data set. An overview of the results is presented in Table 1.

There were no significant between-subject differences for the dependent variables. However, a significant within-subject difference was found regarding the development of trust towards ROBERT. Therefore, the difference between trust ratings before and after completion of the project was assessed using a Wilcoxon-Signed-Ranks test (initial trust was measured with propensity towards trust in autonomous systems questionnaire). In comparison to the baseline group, which showed no difference ($Z = -0.39,\ p = 0.71$), the group with the proactive assistant showed a significantly higher trust building ($Z = -2.58,\ p < 0.01$). The results are visualised in Fig. 4.

Significant gender differences regarding experience in DIY-tools and speech assistants were found. Chi-Squared tests revealed that females had far less experience in the usage of two out of four questioned electric tools (percussion drill, $\chi^2 = 4.63,\ p = 0.03$; electric jigsaw, $\chi^2 = 5.10,\ p = 0.02$) than men. Besides, females had less experience in two out of three speech assistants (Google Assistant, $\chi^2 = 6.52,\ p = 0.01$; Alexa, $\chi^2 = 4.51,\ p = 0.03$). Therefore, gender-based differences regarding proactive meta-dialogue were considered. Females rated the proactive meta-dialogue with ROBERT significantly higher than interacting with the baseline variant for the categories satisfaction ($U = 8.00,\ Z = -2.50,\ p = 0.01$) and speed ($U = 13.00,\ Z = -2.00,\ p = 0.046$). Furthermore, there was a notable higher rating for acceptance of proactive behaviour by females ($U = 14.50,\ Z = -1.81,\ p = 0.07$). Males evaluated the personal attachment towards the baseline assistant notably higher than the proactive version of ROBERT ($U = 4.00,\ Z = -1.49,\ p = 0.07$). Regarding the age of the participants no significant differences were found.

7 DISCUSSION

The study revealed differences between the proactive and the baseline condition. Particularly, it could be shown that the integration of proactive-meta dialogue leads to a significantly better establishment of trust between the user and ROBERT as compared to the non-proactive baseline (H1 verified). This result is quite intuitive, as a system that actively engages with the user and tries to participate in the project should be perceived as a more trustworthy assistant. Especially for novices, to which the entry to a new topic can be quite challenging, it seems to be beneficial to have a more natural and social assistant than a rigid, non-communicative system. Therefore, the socialising purpose of reflective dialogue is useful, as it establishes a communication and acknowledges the progress made by the user in order to foster the user's self-appraisal. Besides the positive effect of reflective dialogue on the user's learning in

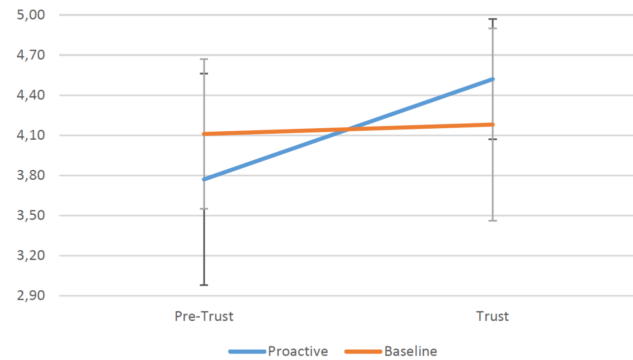

Figure 4: Depiction of the results for the development of trust towards ROBERT depending on the study condition. Mean values and standard deviations from the user ratings on a 5-point Likert scale are provided.

intelligent tutoring systems [35], this work provides evidence that it can also be beneficial for establishing trust in CAs.

Even though there were no significant group differences found regarding the overall measurements of trust and its components (H2 rejected), a trend can be seen that timely proactivity enhances trust towards assistants. Means for overall trust and cognitive-based trust (competence, reliability, and predictability) were all higher for the proactive condition than for the baseline condition. This is in line with related work that shows that low- to medium-level proactivity positively affects a system's perceived competence and reliability [40]. For affect-based trust an opposite tendency is observable. Both variables, personal attachment and faith, were rated higher for the baseline condition. This may be explained by noting that users were more familiar with reactive systems. However, the overall score of these variables was rather low for both conditions. For fostering a better affect-based trust, a long-term relationship with the user, where he or she could emotionally bond with the system, is necessary. Furthermore, the standard deviations were quite high for the affect-based variables. This could be a sign that proactive behaviour has a very individual impact on the user's affect towards assistants.

A positive trend of proactive-meta dialogue in CAs was also noticeable considering the acceptance of and user experience with ROBERT. Means for acceptance, satisfaction, and habitability were higher for the proactive condition. This is similar to the results of a Wizard-of-Oz study reported by Peng et al. [59], where a medium-proactive robot was perceived as more appropriate and helpful in decision-making tasks. However, the differences found in our work are not significant (H3 rejected). Nonetheless, the effect of proactive dialogue strategies on the user's acceptance and satisfaction should be further examined, as proactive behaviour shows much potential. Interestingly, there was no difference between the two conditions regarding the rating of annoyance, as proactive intervention can be perceived as obtrusive. Therefore, the possibility to ignore the system-initiated dialogues, i.e. to let the user have the control over the interaction, was appropriate in this scenario. As proactive dialogue requires the attention of the user, it was no surprise that

	Dialogue Strategy					
	Proactive *M (SD)*			**Baseline** *M (SD)*		
	male	female	overall	male	female	overall
Trust*	4.49 (.33)	4.55 (.54)	4.52 (.45)	4.11 (1.01)	4.22 (.58)	4.18 (.72)
Acceptance*	3.71 (.94)	4.44 (.60)	4.14 (.81)	3.67 (.83)	3.88 (.73)	3.80 (.74)
Reliability	4.68 (1.65)	5.60 (.98)	5.22 (1.32)	4.76 (1.82)	5.18 (.82)	5.03 (1.22)
Competence	4.88 (1.40)	5.46 (1.14)	5.22 (1.23)	4.72 (2.11)	4.80 (.81)	4.77 (1.33)
Predictability	5.90 (1.21)	6.07 (1.05)	6.00 (1.07)	5.40 (1.42)	6.17 (.63)	5.89 (1.00)
Personal Attachment	2.28 (1.60)	3.23 (1.60)	2.83 (1.60)	4.56 (2.20)	2.93 (1.20)	3.51 (1.74)
Faith	3.36 (2.17)	4.49 (1.17)	4.02 (1.67)	4.40 (1.64)	4.82 (1.01)	4.67 (1.22)
Satisfaction	4.93 (1.50)	6.32 (.45)	5.74 (1.20)	5.44 (1.84)	5.14 (1.04)	5.25 (1.32)
Cognitive Demand	4.96 (.90)	5.46 (.87)	5.25 (.88)	5.68 (1.21)	5.27 (1.13)	5.41 (1.13)
Annoyance	4.04 (.68)	5.20 (1.25)	4.72 (1.17)	4.92 (.93)	4.56 (1.08)	4.69 (1.01)
Habitability	5.00 (.77)	5.21 (1.64)	5.13 (1.30)	5.00 (1.76)	4.92 (1.41)	4.95 (1.47)
Speed	5.80 (1.10)	6.29 (1.29)	6.08 (1.18)	5.60 (1.52)	5.28 (1.18)	5.39 (1.26)

Table 1: Descriptive statistics of the measured dependent variables with reference to the the dialogue strategies. Results for cognitive demand and annoyance are inverted (the higher, the better). *Trust and acceptance measured on 5-point Likert scales.

the non-proactive variant of ROBERT was rated better regarding cognitive demand.

Furthermore, gender-dependent effects of proactive meta-dialogue were found. There was a significant difference between proactive and baseline condition for satisfaction and system speed rated by females. Additionally, females tended to accept the proactive assistant more. These gender effects seemed to be due to differences in the experience with DIY-tools and speech assistants between males and females. The more experienced males showed the tendency to prefer working with the non-proactive assistant, as can be seen in the ratings of reliability, personal attachment, satisfaction, and annoyance. This could be explained by the hypothesis that men might have felt patronised by the active assistance and found that the help offered was unnecessary, as they could work independently based on their tool knowledge. Contrarily, females tended to welcome the more communicative guidance by ROBERT, as they required a higher level of cooperation. Another explanation could be that a female system voice was used. Hence, it could be that males did not feel comfortable working under female guidance on a stereotypical male task, as opposed to females. Gender stereotypes have shown to play an important role in spoken human-robot interaction [38]. Therefore, further investigation on their effect on the perception of proactive behaviour could be an interesting research topic.

Comparing the strengths and weaknesses of the study design, the advantages of the setup formed a highly realistic test scenario with a sophisticated CA. A disadvantage was the quite low number of participants. Using a higher number could have provided more comparable results between the two conditions. The complex study setup was also a drawback. Besides the manipulation of proactive assistant behaviour, also tests regarding the user-adaptive planning functionality of ROBERT were conducted simultaneously. Studying these effects separately could have led to more significant results, but was impracticable due to duration and cost of the study. Besides, the quality of the speech synthesis was troublesome as several participants of the study reported the synthetic voice to be unnatural and even annoying. Another problem was that speech had to be activated by "push-to-talk" for technical reasons. Even though users were instructed to use this kind of voice activation in the tutorial, they found this to be cumbersome and annoying.

8 CONCLUSION

In this paper, the integration of proactive meta-dialogue strategies in the companion system ROBERT was presented. The effect of proactive behaviour on the user's perceived trust, acceptance, and user experiences was evaluated in a realistic test scenario. Therefore, a proactive version of ROBERT was compared to a non-proactive baseline variant. The results showed that proactive meta-dialogue was able to build user trust significantly better than a solely reactive system. Additionally, gender-specific differences for the perception of proactive meta-dialogue were found. However, differences between overall scores were not significant. Although the scores showed a tendency towards a positive effect of proactive dialogue on trust, acceptance, and user experience, further experiments are necessary to validate these tendencies. Therefore, studies in other application domains need to be considered as well. Additionally, investigation of proactive behaviour in the background of gender bias could form a new interesting research topic.

ACKNOWLEDGMENTS

This work is done within the technology transfer project “Do it yourself, but not alone: *Companion*-Technology for Home Improvement” of the Transregional Collaborative Research Centre SFB/TRR 62 “*Companion*-Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG). The industrial project partner is the Corporate Research Sector of the Robert Bosch GmbH. We thank MTO Psychologische Forschung und Beratung GmbH for facilitating the reported experiments.

REFERENCES

[1] IfD Allensbach. 2019. Allensbacher Markt- und Werbetraeger-Analyse – AWA 2019, zitiert nach de.statista.com. https://de.statista.com/statistik/daten/studie/170918/umfrage/interesse-an-heimwerken-und-do-it-yourself/ Retrieved on 07.04.2020 at 18.37.

[2] John R Anderson, C Franklin Boyle, and Brian J Reiser. 1985. Intelligent tutoring systems. *Science* 228, 4698 (1985), 456–462.

[3] Jimmy Baraglia, Maya Cakmak, Yukie Nagai, Rajesh Rao, and Minoru Asada. 2016. Initiative in robot assistance during collaborative task execution. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI 2016)*. IEEE Press, 67–74.

[4] Jenay M Beer, Arthur D Fisk, and Wendy A Rogers. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction* 3, 2 (2014), 74–99.

[5] Gregor Behnke, Pascal Bercher, Susanne Biundo, Birte Glimm, Denis Ponomaryov, and Marvin Schiller. 2015. Integrating Ontologies and Planning for Cognitive Systems. In *Proceedings of the 28th International Workshop on Description Logics (DL 2015)*. CEUR Workshop Proceedings, 338–350.

[6] Gregor Behnke, Pascal Bercher, Matthias Kraus, Marvin Schiller, Kristof Mickeleit, Timo Häge, Michael Dorna, Michael Dambier, Wolfgang Minker, Birte Glimm, and Susanne Biundo. 2020. New Developments for Robert – Assisting Novice Users Even Better in DIY Projects. In *Proc. of the 30th Int. Conf. on Autom. Plan. and Sched. (ICAPS 2020)*. AAAI Press, 343–347.

[7] Gregor Behnke, Daniel Höller, and Susanne Biundo. 2018. totSAT – Totally-Ordered Hierarchical Planning through SAT. In *Proc. of the 32nd AAAI Conf. on AI (AAAI 2018)*. AAAI Press, 6110–6118.

[8] Gregor Behnke, Daniel Höller, and Susanne Biundo. 2018. Tracking branches in trees – A propositional encoding for solving partially-ordered HTN planning problems. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI 2018)*. IEEE, 73–80.

[9] Gregor Behnke, Daniel Höller, and Susanne Biundo. 2019. Bringing order to chaos – A compact representation of partial order in SAT-based HTN planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. AAAI, AAAI Press, 7520–7529.

[10] Gregor Behnke, Daniel Höller, and Susanne Biundo. 2019. Finding Optimal Solutions in HTN Planning – A SAT-based Approach. In *Proc. of the 28th Int. Joint Conf. on AI (IJCAI 2019)*. IJCAI, 5500–5508. https://doi.org/10.24963/ijcai.2019/764

[11] Gregor Behnke, Denis Ponomaryov, Marvin Schiller, Pascal Bercher, Florian Nothdurft, Birte Glimm, and Susanne Biundo. 2015. Coherence Across Components in Cognitive Systems – One Ontology to Rule Them All. In *Proc. of the 24th Int. Joint Conf. on AI (IJCAI 2015)*. IJCAI, 1442–1449.

[12] Gregor Behnke, Marvin Schiller, Matthias Kraus, Pascal Bercher, Mario Schmautz, Michael Dorna, Michael Dambier, Wolfgang Minker, Birte Glimm, and Susanne Biundo. 2019. Alice in DIY wonderland or: Instructing novice users on how to use tools in DIY projects. *AI Communications* 32, 1 (2019), 31–57.

[13] Gregor Behnke, Marvin Schiller, Matthias Kraus, Pascal Bercher, Mario Schmautz, Michael Dorna, Wolfgang Minker, Birte Glimm, and Susanne Biundo. 2018. Instructing Novice Users on How to Use Tools in DIY Projects. In *Proc. of the 27th Int. Joint Conf. on AI and the 23rd Europ. Conf. on AI (IJCAI-ECAI 2018)*. IJCAI, 5805–5807.

[14] Pascal Bercher, Ron Alford, and Daniel Höller. 2019. A Survey on Hierarchical Planning – One Abstract Idea, Many Concrete Realizations. In *Proc. of the 28th Int. Joint Conf. on AI (IJCAI 2019)*. IJCAI, 6267–6275.

[15] Pascal Bercher, Susanne Biundo, Thomas Geier, Thilo Hoernle, Florian Nothdurft, Felix Richter, and Bernd Schattenberg. 2014. Plan, repair, execute, explain – how planning helps to assemble your home theater. In *Twenty-Fourth International Conference on Automated Planning and Scheduling (ICAPS 2014)*. AAAI Press, 386–394.

[16] Susanne Biundo and Andreas Wendemuth. 2016. Companion-technology for cognitive technical systems. *KI – Künstliche Intelligenz* 30, 1 (2016), 71–75.

[17] Christine Cheepen. 1988. *The predictability of informal conversation.* Pinter Pub Ltd.

[18] Wei-Yi Cheng, Alf Scotland, Florian Lipsmeier, Timothy Kilchenmann, Liping Jin, Jens Schjodt-Eriksen, Detlef Wolf, Yan-Ping Zhang-Schaerer, Ignacio Fernandez Garcia, Juliane Siebourg-Polster, et al. 2017. Human activity recognition from sensor-based large-scale continuous monitoring of Parkinson's disease patients. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE 2017)*. IEEE, 249–250.

[19] Robin I M Dunbar. 1998. *Grooming, gossip, and the evolution of language.* Harvard University Press.

[20] Amos Freedy, Ewart DeVisser, Gershon Weltman, and Nicole Coeyman. 2007. Measurement of trust in human-robot collaboration. In *2007 International Symposium on Collaborative Technologies and Systems*. IEEE, 106–114.

[21] Lone Frølund and Jan Nielsen. 2009. The reflective meta-dialogue in psychodynamic supervision. *Nordic Psychology* 61, 4 (2009), 85–105.

[22] Arthur C Graesser, Kurt VanLehn, Carolyn P Rosé, Pamela W Jordan, and Derek Harter. 2001. Intelligent tutoring systems with conversational dialogue. *AI magazine* 22, 4 (2001), 39–52.

[23] Adam M Grant and Susan J Ashford. 2008. The dynamics of proactivity at work. *Research in organizational behavior* 28 (2008), 3–34.

[24] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, convolutional, and recurrent models for human activity recognition using wearables. In *Proc. of the 25th Int. Joint Conf. on AI (IJCAI 2016)*. 1533–1540.

[25] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527.

[26] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.

[27] Kate S Hone and Robert Graham. 2000. Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering* 6, 3-4 (2000), 287–303.

[28] Chien-Ming Huang and Bilge Mutlu. 2016. Anticipatory robot control for efficient human-robot collaboration. In *The eleventh ACM/IEEE international conference on human robot interaction (HRI 2016)*. IEEE Press, 83–90.

[29] David Hustadt. 2019. *Praktikumsbericht WS 18/19.* Internship Report. Hochschule der Medien Stuttgart, in cooperation with Robert Bosch GmbH.

[30] Jeffrey L Jenkins, Bonnie Brinton Anderson, Anthony Vance, C Brock Kirwan, and David Eargle. 2016. More harm than good? How messages that interrupt can make us vulnerable. *Information Systems Research* 27, 4 (2016), 880–896.

[31] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71.

[32] Katja Karrer, Charlotte Glaser, Caroline Clemens, and Carmen Bruder. 2009. Technikaffinität erfassen – der Fragebogen TA-EG. In *Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme*. VDI Verlag GmbH, 196–201.

[33] Sandra Katz, Patricia Albacete, and Pamela Jordan. 2016. Do summaries support learning from post-problem reflective dialogues?. In *Intelligent Tutoring Systems*. Springer, 519–520.

[34] Sandra Katz, John Connelly, and Christine Wilson. 2007. Out of the lab and into the classroom: An evaluation of reflective dialogue in Andes. In *Artificial Intelligence in Education (AIED 2017)*. IOS Press, 425–432.

[35] Sandra Katz, Gabriel O'Donnell, and Heather Kay. 2000. An Approach to Analyzing the Role and Structure of Reflective Dialogue. *International Journal of Artificial Intelligence in Education* 11 (2000), 320–343.

[36] Alice Kerlyl, Phil Hall, and Susan Bull. 2006. Bringing chatbots into education: Towards natural language negotiation of open learner models. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, 179–192.

[37] Stefan Kopp, Lars Gesellensetter, Nicole C Krämer, and Ipke Wachsmuth. 2005. A conversational agent as museum guide – design and evaluation of a real-world application. In *International workshop on intelligent virtual agents*. Springer, 329–343.

[38] Matthias Kraus, Johannes Kraus, Martin Baumann, and Wolfgang Minker. 2018. Effects of Gender Stereotypes on Trust and Likability in Spoken Human-Robot Interaction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. ELRA, 112–118.

[39] Matthias Kraus, Marvin Schiller, Gregor Behnke, Pascal Bercher, Susanne Biundo, Birte Glimm, and Wolfgang Minker. 2019. A Multimodal Dialogue Framework for Cloud-Based Companion Systems. In *Proc. of the 9th Int. Work. on Spoken Dialogue Systems Technology (IWSDS 2019)*. Springer, 405–410.

[40] Matthias Kraus, Nicolas Wagner, and Wolfgang Minker. 2020. Effects of Proactive Dialogue Strategies on Human-Computer Trust. In *Proceedings of the 28th Conference on User Modeling, Adaptation and Personalization (UMAP 2020)* (Genoa, Italy). Association for Computing Machinery, 107–116.

[41] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258.

[42] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[43] Jinna Lei, Xiaofeng Ren, and Dieter Fox. 2012. Fine-grained kitchen activity recognition using RGB-D. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (Ubicomp 2012.* Association for Computing Machinery, 208–211.

[44] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *11th Australasian Conference on Information Systems (ACIS 2000)*. Information Systems Management Research Centre, Queensland University of Technology. Retrieved on 27.08.2020 from http://acis.aaisnet.org/archive.php.

[45] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[46] Daniel C McFarlane and Kara A Latorella. 2002. The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction* 17, 1 (2002), 1–61.

[47] Paul McIntosh. 2010. *Action research and reflective practice: Creative and visual methods to facilitate reflection and learning.* Routledge.

[48] Michael Frederick McTear, Zoraida Callejas, and David Griol. 2016. *The conversational interface.* Springer.

[49] Stephanie M Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human factors* 55, 3 (2013), 520–534.

[50] Bonnie M Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (1996), 429–460.

[51] Karen Myers and Neil Yorke-Smith. 2007. Proactive behavior of a personal assistive agent. In *Proceedings of the AAMAS Workshop on Metareasoning in Agent-Based Systems, Honolulu, HI.* ACM, 31–45.

[52] Alexander Neumann, Christof Elbrechter, Nadine Pfeiffer-Leßmann, Risto Kõiva, Birte Carlmeyer, Stefan Rüther, Michael Schade, André Ückermann, Sven Wachsmuth, and Helge J Ritter. 2017. "KogniChef": A Cognitive Cooking Assistant. *KI-Künstliche Intelligenz* 31, 3 (2017), 273–281.

[53] Florian Nothdurft, Gregor Behnke, Pascal Bercher, Susanne Biundo, and Wolfgang Minker. 2015. The interplay of user-centered dialog systems and AI planning. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2015).* ACL, 344–353.

[54] Florian Nothdurft, Felix Richter, and Wolfgang Minker. 2014. Probabilistic human-computer trust handling. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2014).* ACL, 51–59.

[55] Florian Nothdurft, Stefan Ultes, and Wolfgang Minker. 2015. Finding appropriate interaction strategies for proactive dialogue systems-an open quest. In *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication, August 6-8, 2014, Tartu, Estonia.* Linköping University Electronic Press, 73–80.

[56] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.

[57] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 2008. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of cognitive engineering and decision making* 2, 2 (2008), 140–160.

[58] Sharon K Parker, Helen M Williams, and Nick Turner. 2006. Modeling the antecedents of proactive behavior at work. *Journal of applied psychology* 91, 3 (2006), 636–652.

[59] Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and Evaluation of Service Robot's Proactivity in Decision-Making Support Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM, paper 98.

[60] Heather Pon-Barry, Brady Clark, Karl Schultz, Elizabeth Owen Bratt, Stanley Peters, and David Haley. 2005. Contextualizing reflective dialogue in a spoken conversational tutor. *Journal of Educational Technology & Society* 8, 4 (2005), 42–51.

[61] Joseph Rafferty, Chris D Nugent, Jun Liu, and Liming Chen. 2017. From activity recognition to intention recognition for assisted living within smart homes. *IEEE Transactions on Human-Machine Systems* 47, 3 (2017), 368–379.

[62] Anand S Rao and Michael P Georgeff. 1995. BDI agents: from theory to practice.. In *Proceedings of the First International Conference on Multiagent System (ICMAS-95).* MIT Press, 312–319.

[63] Pei-Luen Rau, Ye Li, and Jun Liu. 2013. Effects of a Social Robot's Autonomy and Group Orientation on Human Decision-Making. *Advances in Human-Computer Interaction* 2013 (2013), Article ID 263721, 13 pages. https://doi.org/10.1155/2013/263721

[64] Marvin Schiller, Gregor Behnke, Mario Schmautz, Pascal Bercher, Matthias Kraus, Michael Dorna, Wolfgang Minker, Birte Glimm, and Susanne Biundo. 2017. A Paradigm for Coupling Procedural and Conceptual Knowledge in Companion Systems. In *Proc. of the 2nd Int. Conf. on Companion Tech. (ICCT 2017).* IEEE, 1–6. https://doi.org/10.1109/COMPANION.2017.8287072

[65] Klaus P Schneider. 1988. *Small talk: Analyzing phatic discourse.* Sprachwissenschaftliche Reihe, Vol. 1. Hitzeroth.

[66] Choong-Nyoung Seon, Harksoo Kim, and Jungyun Seo. 2012. A statistical prediction model of speakers' intentions using multi-level features in a goal-oriented dialog system. *Pattern Recognition Letters* 33, 10 (2012), 1397–1404.

[67] Thomas B Sheridan and William L Verplank. 1978. *Human and computer control of undersea teleoperators.* Technical Report. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

[68] Xing Su, Hanghang Tong, and Ping Ji. 2014. Activity recognition with smartphone sensors. *Tsinghua science and technology* 19, 3 (2014), 235–249.

[69] David R Traum, Carl F Andersen, Waiyian Chong, Darsana Josyula, Yoshi Okamoto, Khemdut Purang, Michael O'Donovan-Anderson, and Don Perlis. 1999. Representations of dialogue state for domain and task independent meta-dialogue. *Electronic Transactions on Artificial Intelligence* 3, Section D (1999), 125–152.

[70] Jinke D Van Der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation research. Part C, Emerging technologies* 5, 1 (1997), 1–10.

[71] Nicolas Wagner, Matthias Kraus, Niklas Rach, and Wolfgang Minker. 2019. How to address humans: System barge-in in multi-user HRI. In *Proceedings of the 10th International Workshop on Spoken Dialog Systems Technology (IWSDS 2019)* (Siracusa (Sicily, Italy)).

[72] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2019. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters* 119 (2019), 3–11.

[73] Yu Zhang, Vignesh Narayanan, Tathagata Chakraborti, and Subbarao Kambhampati. 2015. A human factors analysis of proactive support in human-robot teaming. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015).* IEEE, 3586–3593.